PRESIDENCY UNIVERSITY

DEPARTMENT OF LIFE SCIENCES

PhD Thesis Synopsis

Title- **Characterization of Aureochromes: the Blue Light Responsive LOV Photoreceptor-Transcription Factors from Photosynthetic Marine Algae**

Name of the Candidate	**MADHURIMA KHAMARU**
Registration Number	**R-19RS218210210**
Date of Registration	**23.02.2021**
Name of Supervisor	**Dr. Devrani Mitra**

Name of the Candidate: MADHURIMA KHAMARU

Name of Supervisor: Dr. DEVRANI MITRA

Title of the Thesis: Characterization of Aureochromes: the Blue Light Responsive LOV Photoreceptor-Transcription Factors from Photosynthetic Marine Algae

Registration No.: R-19RS218210210; Date: 23.02.2021

Enrolment Number: PU/PHD/DLS/2019/079; Date: 02.01.2020

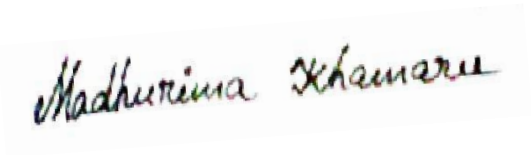
Place of work: Presidency University, Kolkata

Residential Address: 19, N.C. Pakrashi Lane; P.O.- Rishra; Dist.– Hooghly, PIN-712248, West Bengal, India

Corresponding Address: Same as Residential Address

Email: madhurima.khamaru15@gmail.com

Phone Number: 9432873003/ 9477000234



Signature of Research Scholar **Signature of Supervisor**

CONTENTS

TITLE	PAGE NUMBERS
Abstract	4
Introduction	4-5
Objectives	6
Methodology	6-12
Results &Discussion	12-49
Conclusion	49-50
Significance of the study	50-51
Future Direction	51-52
References	52-56
Plagiarism Certificate	57
Additional Documents	58-60

ABBREVIATION

BL – Blue light, LOV- Light-Oxygen-Voltage, TF-Transcription factor BLUF - Blue-light sensors using flavin adenine dinucleotide, bZIP- basic leucine zipper, STK- Ser/Thr kinase, HTH- helix-turn-helix, HK- Histidine-kinase, Es- *Ectocarpus siliculosus,* Aureo- Aureochrome, MD-Molecular Dynamics simulations, RIG- Residue Interaction Graphs, EC- Eigenvector centrality, ΔE_{SP}- specific difference of Energy of a Graph, P_{av}- Average Participation Coefficient, B_r- Relative change in Betweenness Centrality, KL- KL Divergence, I_r- Relative change in Information Centrality, SDS-PAGE- Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis, EMSA-Electrophoretic mobility shift assays, MSA- Multiple Sequence Alignment, MCS-Multiple cloning site

ABSTRACT

Photoreceptors are the light sensitive proteins that absorb photon from the visible region of electromagnetic spectrum. Ten different types of photoreceptors can be classified based on their associated sensor like- Light-Oxygen-Voltage/LOV, Xanthopsin, Blue-light-utilizing-FAD(BLUF), Cryptochrome, Phytochrome, Rhodopsin. Aureochromes, the potential blue-light (BL) responsive LOV photoreceptors of photosynthetic marine stramenopiles other than cryptochromes, regulate diverse biological activities. These BL dependent transcription-factors(TFs) exhibiting unique association of DNA binding basic-leucine-zipper(bZIP) domain with LOV sensor, carry out BL dependent gene expression. Few homologs of aureochromes have been characterized since its discovery in 2007[1]. The species-specific physiological functions of aureochromes are the result of subtle differences in amino acid sequences and oligomerization properties. This study therefore aims for structure-function characterization of *Ectocarpus siliculosus* aureochromes with further exploration of their dimerization potential. Additionally, *Thallasiosira oceanica* aureochrome is engineered to generate synthetic variants suitable for applications in plant/algal optogenetics.

INTRODUCTION

Light, the electromagnetic wave containing different wavelength associated with different energies is essential for the light responsive signal transduction cascades of autotrophic organisms. Since the wavelength of BL shows maximum penetrance in sea water, transcription of most of the genes

important for growth and developmental processes of photosynthetic marine algae are BL-regulated. Aureochromes are the unique BL (390-500nm) sensing photoreceptors with TF like activity found in the 'photosynthetic stramenopiles'[23]/algal group-heterokontophyta[4]. The BL sensor in aureochromes belongs to LOV category – the major group among the ten different classes of photoreceptors. LOV photoreceptors[5] can be associated with a variety of effectors[6] in a modular architecture, such as- STK in phototropin, HTH in EL222 to HK in LOV-STAS. But, aureochrome is the only example where bZIP constitutes the effector domain. Aureochrome being the only light responsive bZIP, merits thorough characterization. *Ectocarpus siliculosus* of phaeophyceae possess five paralogs of aureochrome. The habitat of these algae in the rocky coastal ecosystem is prone to light fluctuation accompanied by biotic stress. Aureochromes, present in multiple copies, are thus supposed to perform BL mediated diverse gene regulatory activities through its basic leucine zipper (bZIPs)[7,8]. In fact, characterization of the LOV and bZIP domains of all its aureochromes is important for understanding the differences among the subtypes as it possesses the highest number of aureochrome paralogs. BL-triggered aureochrome mediated gene regulatory network is better to be deciphered in *Ectocarpus siliculosus*[9] for its small genome size and feasibility of genetic modification. bZIPs binds to DNA essentially as dimers. Heterodimerization leads to selection of substrates from a wide variety of substrate DNA repertoires for the regulation of diverse biological pathways. It is, therefore, crucial to determine the partnering specificity of aureochrome-bZIPs and exploring the diversity of dimers. Aureochrome is an ideal natural transcription factor to be used for the development of an ever-growing suite of opto-genetic tools for its prompt light regulated DNA binding activity and eukaryote-like inverse effector-sensor domain topology. The diatom species *Thallasiosira oceanica* play a vital role in carbon cycling in the marine and freshwater ecosystems as the key primary producer. The aureochrome from *Thallasiosira oceanica* exhibits ideal photochemical kinetics and light regulated DNA binding ability with nearly four-fold difference in binding affinity in the presence of light[10]. So, engineering of this aureochrome using directed evolutionary approach has been attempted for the generation of its superior synthetic variety to be used in plant/algal optogenetics as a reversible photo switch.

OBJECTIVES

1. Biochemical, biophysical and *in-silico* characterization of aureochromes from the photosynthetic marine stramenopile e.g. *Ectocarpus siliculosus*
2. Understanding combinatorial diversity of aureochrome bZIPs through genome wide screening in *Ectocarpous sp.*
3. Exploring the potential of aureochromes in plant/algal optogenetics via synthetic studies

METHODOLOGY

- ***In-silico*-**

Sequence Retrieval and Analysis:

The proteins whose crystal structures have been solved were obtained from PDB[11]. Other amino acid sequences including all the Aurechromes, some bZIP proteins and the LOV domain containing proteins were collected from the protein sequence database- UniProt[12]. Domain analysis of the retrieved protein sequences was done by NCBI-BLAST[13]. To determine the sequential differences among the proteins, extensive multiple sequence alignment was performed with the use of Clustal Omega[14].

Phylogenetic Studies:

Phylogenetic trees were made by neighbor joining (NJ) method in MEGAX[15] maintaining 1000 bootstrap replications for each node. Alignment of sequences were done using Clustral W[16].

Protein Modelling, Molecular Docking and *in silico* interaction study:

The proteins were modeled by SWISS-MODEL[17] using the appropriate template. The best models selected on the basis of Ramachandran parameters were energy minimized using Chimera[18]. The energy minimized structures were subjected to molecular docking using NPDock[19]. The docked structures with the highest docking confidence score were analyzed by PDBePISA[20] and visualized using Ligplot+ and Pymol[21].

Molecular Dynamics Simulation:

MD simulations were performed using GROMACS (v2020.1 and 2021)[22]. The AMBER99SB-ILDN/charm27 all-atom force field is applied and solvated with TIP3P/ SPC water model in a 1.5

nm^3 cubic lattice. Systems were neutralized by the addition of chloride (Cl^-) ions. Energy minimization was done by steepest descent followed by the conjugate gradient algorithm (50,000 steps each) to resolve steric clashes. The stability of the systems was ensured by maintaining a negative potential energy (Epot) before the equilibration run. NVT (100 ps) and NPT (100 ps) ensembles were conducted with position restraints on non-water residues and H-bond constraints. Berendsen thermostat was used to maintain the temperature at 300 K. Particle Mesh Ewald (PME) handled long-range electrostatics and pressure fluctuations were averaged with a reference of 1 bar (isothermal compressibility: 4.5 x 10^{-5} bar^{-1} for water). Bond restraints (including hydrogens) were applied using LINCS, with an integration time step of 2 fs that updates every 1 ps. The equilibration run was carried out for a time frame of 60ns by three consecutive runs of 20 ns at 300 K. The post-simulation trajectory files were analysed using the command-line tools of GROMACS[23].

Network Analysis:

Overall connectivity in the protein-DNA and protein-protein structures was depicted through the construction of RIGs using Biographs. The protein-DNA complexes of eight representative homo/heterodimeric bZIP co-crystal structures were represented as RIG, where nodes i.e. amino acids/nucleotide residues were linked by edges within a given distance cut off (e.g. 4.5 Å). EC[24] of the constructed RIGs that represents the degree of connectivity of a node with other well-connected nodes of a network, was calculated to capture all the communications at the residue level. Protein-protein dimer stability was calculated through network-theoretic (e.g., ΔE_{SP},P_{av}, B_r) and information-theoretic (e.g., KL and Ir) approaches. ΔE_{SP} accords a tool from the perspective of topological analyses of networks to study the stability of a protein, Pav predicts stability of the dimer in terms of higher inter monomeric connections. KL assesses dimer stability by calculating the dissimilarity between two probability distributions, B_r calculates[25] available internodal shortest paths to predict dimer stability and I_r measures[24] holistic changes in paths of a graph upon dimerization.

\

- ***In-Vitro-***

Cloning of Synthetic genes:

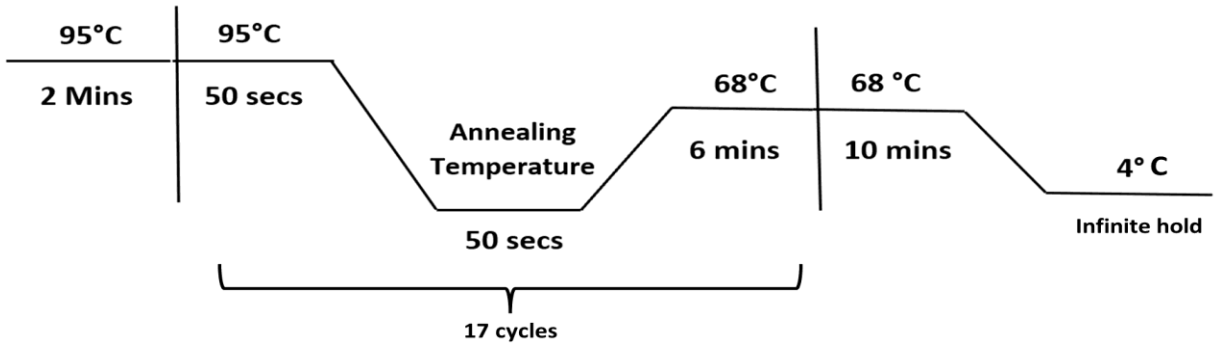
Objective-1: The purchased codon optimized synthetic genes (Aureochrome-1, uniport ID- D7FMH8; Aureochrome-5, uniport ID- D7FQI9) in the pUC-57 plasmids were first of all amplified to incorporate the NheI/SacI restriction sites into them. The expression vector pET28a and the amplified DNA harboring the NheI/SacI restriction site were then double digested with the respective enzymes from NewEnglandBioLab followed by the ligation of the restriction digested gene into the T7promoter containing pET28a vector at the MCS right after the poly-histidine tag.

Objective-3: The error-prone PCR products were subjected to restriction digestion by restriction enzymes -BamHI and SacI (New England BioLabs). Every double-digested DNA fragment possessing the BamHI/SacI restriction site was then ligated to pET28a vector at the BamHI/Sac1 restriction sites within the MCS.

The ligated constructs were transformed into the DH5α strain of *Escherichia coli* competent cells for multiplication. The screening of the cloned vectors was done through PCR of the colonies obtained in transformation followed by DNA sequencing.

Site Directed Mutagenesis:

Oligonucleotides of 10-30 nucleotide length are designed as the mutational primers after checking their melting temperatures using online tool- OligoCalc for the incorporation of a desired amino acid at a specific site in the DNA by substitution of the original one. Template, forward and reverse primers, Phusion GC 5X buffer, dNTP, NuF, Phusion polymer and DMSO were used to prepare thePCR-mix.

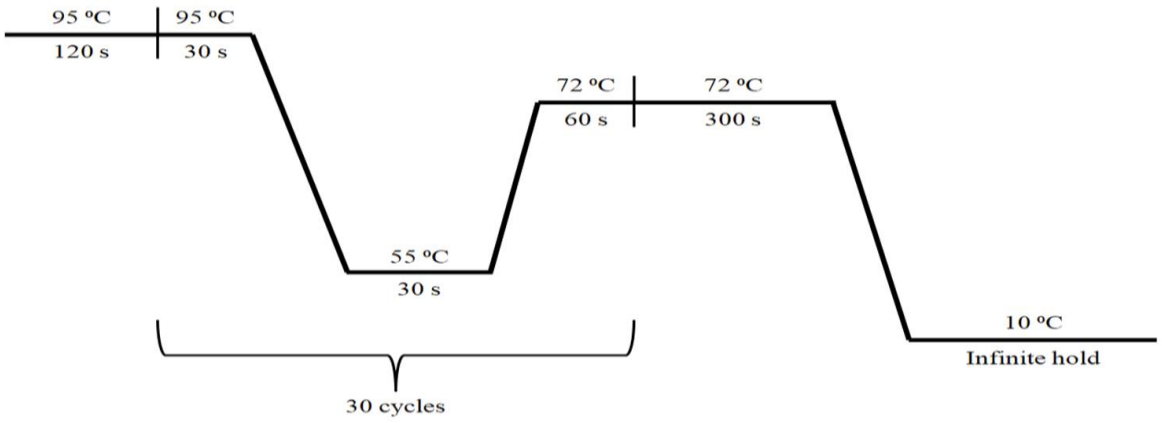


After the run, the PCR products were subjected to DPN1 digestion for 4 hours for the removal of parental strand. The DPN1 digested product was then transformed into the DH5α strain of *E. coli.* Small cultures were grown overnight in the incubator at 37˚C on shaking at 150 r.p.m. Plasmids

were isolated from the overnight grown bacterial culture by Promega plasmid isolation kit. Confirmation of the mutagenesis was done by sequencing of the plasmids.

Error Prone PCR:

Different concentrations of magnesium chloride ($MgCl_2$) and/or manganese chloride ($MnCl_2$) in various combinations were used in the PCR mixture to make it error-prone[26] during the amplification so that random point mutations get introduced. The 3'-5'exonuclease proofreading activity and the fidelity of DNA polymerase got compromised in the presence of Mn^{2+} and Mg^{2+} as these ions decrease the precision of base selection and alters the base pairing.



The PCR products that gave positive bands in the 1% agarose gel were subjected to restriction digestion followed by ligation for their cloning into the pET expression vector. Mutations were validated by DNA sequencing of the cloned plasmids.

Sequence analysis:

Chromatogram files of the DNA sequencing results obtained from the Applied Biosystems DNA sequencers were analyzed using Chromas 2.6.6 (https://technelysium.com.au/wp/chromas/). ExPASy Translate tool was then used to translate the nucleotide sequences into the protein sequences (https://web.expasy.org/translate/). Protein sequence with the desired open reading frame (ORF), containing the polyhistidine-tag (6xHis-tag) at the N-terminal and starting with methionine, was selected and submitted to BLASTX to check whether the correct sequence has been obtained (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx). Multiple sequence alignment by Clustal Omega server was performed in case of mutants to identify the appropriate mutations (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Protein Over-expression:

The plasmids were transformed into the C43 strain of *E. coli* for protein expression. Colonies obtained from the transformation plate were inoculated into 10 mL autoclaved Luria-Bertani (LB) broth culture containing 10 µL of 50 mg/mL kanamycin (HiMedia) and grown at 37 ºC on shaker at 150 rpm for ≈16–18 hours (overnight) in the incubator-shaker (Thermo Fisher Scientific). Each of the overnight grown small culture was then transferred to 1L of autoclaved LB broth containing 1000 µL of 50 mg/mL kanamycin. Bacterial growth in these large cultures was continued for ≈1.5 hours at 37 ºC and 160 rpm, until the OD at 600nm reaches 0.6. To induce protein overexpression 300 µM isopropylthio-β-galactoside (IPTG) (SRL) was then added and the bacterial cell cultures were grown overnight in the dark at 22 ºC and 120 rpm.

Protein Purification:

The induced bacterial culture was pelleted down in multiple rounds of centrifugation by Thermofisher Scientific Cold Centrifuge at 4 ºC and 8,000 rpm for 10 minutes. The cell pellet was dissolved in lysis buffer (20 mM Tris, pH 8; 50 mM NaCl; 10% glycerol) with lysozyme (SRL), 1.5 µL of protease inhibitor cocktail (Sigma-Aldrich) and a pinch of phenylmethylsulfonyl fluoride (PMSF) (SRL). The dissolved pellet was incubated with ice for ≈15–20 minutes at 4 ºC followed by Sonication (Hielscher Ultrasonics) on ice maintaining 30 seconds on - off cycle at an amplitude of 80 until it turns yellow-brown with a liquidly consistency. The lysed pellet was then centrifuged at 4 ºC and 14,000 rpm for 1 hour in a refrigerated centrifuge with fixed-angle rotor (Thermo Fisher Scientific). The supernatant obtained after centrifugation, was passed through affinity chromatographic (Qiagen) column filled with nickel-nitrilotriacetic acid (Ni-NTA) followed by the passing of wash buffer (20 mM Tris, pH 8; 50 mM NaCl; 10% glycerol; 10 mM imidazole). Protein elution fractions were collected using elution buffer (20 mM Tris, pH 8; 50 mM NaCl; 10% glycerol; 200 mM imidazole). Sephadex desalting column was used for the removal of imidazole from the protein fractions of interest. The desalted protein was then concentrated using Amicon centrifugal filter unit (Merck-Millipore) and kept in -20 ºC for the storage and preservation.

UV-Visible Spectroscopy:

The purified and desalted proteins were placed in quartz cuvette of 1ml with 1cm path length and subjected to UV-visible spectrophotometer (Shimadzu). Absorption spectrum of the proteins was then scanned from 200nm to 750nm to record the absorption peaks of the proteins at 447 nm.

Estimation of protein concentration:

Protein concentrations were determined by applying Beer-Lambert law, and extinction coefficient of 12,500 M^{-1} cm^{-1} from absorbance value recorded at 447 nm using the UV-vis spectrophotometer.

$A = \varepsilon\ c\ l$ [A = Absorbance at 447 nm; ε = Extinction coefficient = 12,500 $M^{-1}cm^{-1}$; c = Concentration of the protein; l = Path length = 1 cm]

Fluorescence Spectroscopy:

The fluorescence intensities of the free FMN and the protein of three optical densities were measured keeping the excitation wavelength at 465nm, emission wavelength at the range 495-520nm, the slit 2nm and the integration time 0.2secs. The respective fluorescence intensity values at 496nm and 522 nm for the protein and free FMN were plotted in Y axis plotting the optical density values at X-axis. The quantum yield was calculated from the slope of the straight line fitted plots by the formula- ΦAUREO5 = ΦFMN (mAUREO5 / mFMN) where ΦFMN is 0.26.

Gel Electrophoresis:

Agarose Gel Electrophoresis- 1% Agarose gel was run at 100 volts for 30 mins and visualized under UV-transilluminator.

SDS Gel Electrophoresis- 15% resolving and 4% stacking gel was resolved at 130 volts for 1 hour and stained with Coomassie Blue for an hour followed by removal of the excess stain by destaining solution containing methanol, acetic acid and water.

Electrophoretic Mobility Shift Assay:

The purchased (IDT) lyophilized HPLC purified 24bp long single-stranded oligonucleotides containing 'Aureo-box' sequence were resuspended in annealing buffer as both the complementary strands in equimolar amounts (10 mM Tris, pH 8.0; 20 mM NaCl) and annealed by heating at 95˚C for 2 mins followed by gradual overnight cooling. The binding buffer (50 mM Tris, pH 8.0; 50 mM NaCl; 1.25 mM MgCl2; 0.01 mg/mL BSA; 20% glycerol) was then added to the double

stranded DNA (dsDNA) to achieve a final concentration of 0.5 μM that was used as the substrate DNA. Proteins were serially diluted to be added to the solution mixture of dsDNA and binding buffer. Nine samples of protein-DNA mixtures were prepared with the varying protein concentrations. Only substrate DNA without any protein was taken as the negative control. The control as well as the samples were incubated on ice under blue light for 45 minutes prior to the commencement of electrophoresis. The electrophoretic run was carried out at 4 ºC under dark and the illuminated condition by Bio-Rad's Mini-PROTEAN system. The samples were resolved at 150 Volts for an hour at 4 ºC. The native gel was stained with the nucleic acid stain SYBR Gold (Thermo Fisher Scientific) for 45 minutes and imaged in the Bio-Rad ChemiDoc MP system after the completion of the electrophoretic run.

RESULTS & DISCUSSION

- **Objective-1:**
  Biochemical, biophysical and *in-silico* characterization of aureochromes from the photosynthetic marine stramenopile e.g. *Ectocarpus siliculosus*

  Search through the protein database-UNIPROT revealed the presence of 26 Aureochrome homologs in the different members of photosynthetic stramenopile [**Table-1**]. The Phylogenetic tree of all the aureochrome homologs clearly pointed out the distinctness of aureochrome-2 and aureochrome-5 from rest of the aureochromes. [**Figure-1.a**]. Most aureochrome sub-types (aureochrome-1, aureochrome-3, aureochrome-4) intermingled under a common line of evolution indicating closeness. While all aureochrome-2 together formed a separate clade under a shared lineage, Aureochrome-5 homologs exhibited a separate lineage. Delineation of the bZIP and LOV of each aureochrome using NCBI-BLAST revealed the absence of bZIP domain in *Es*Aureo5 [**Figure-1.b**]. Phylogenetic clustering of the LOV and bZIP depicted their ancestor-descendant relationship among the aureo homologs [**Figure-1.c, 1.d**].

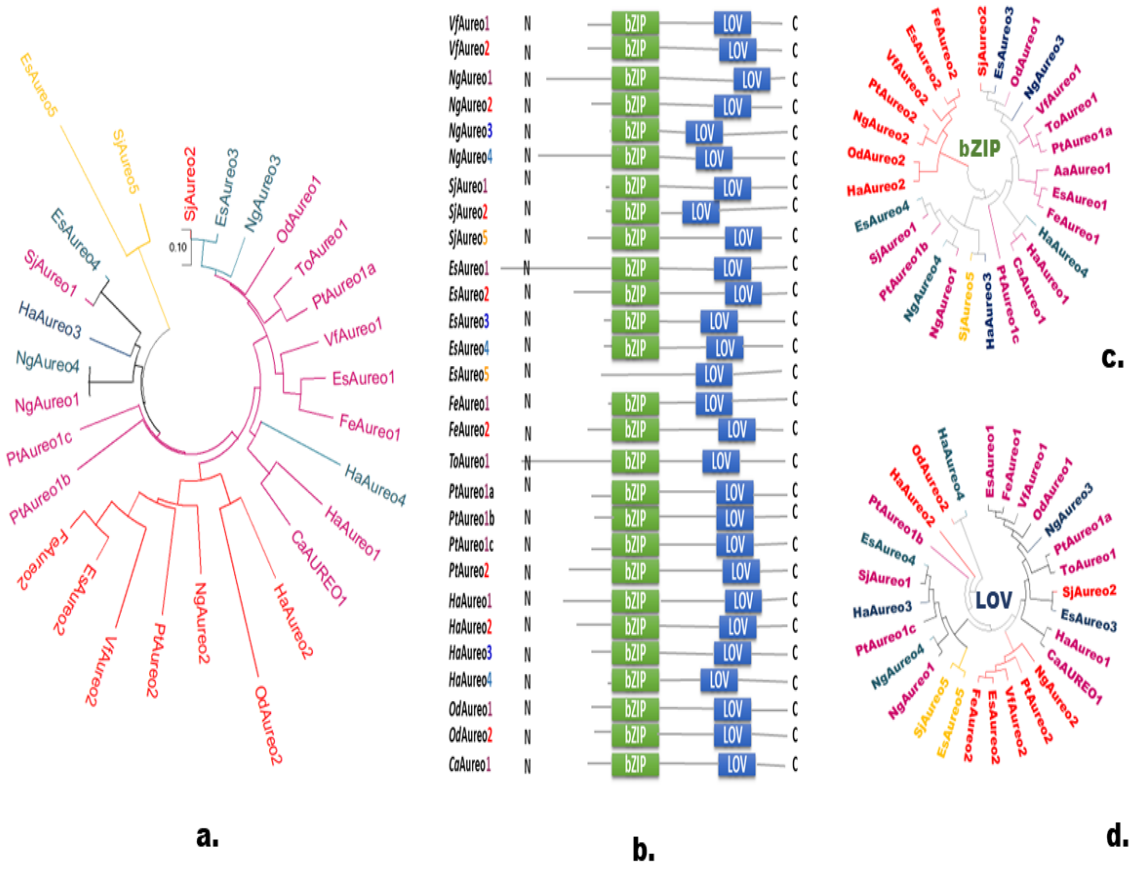
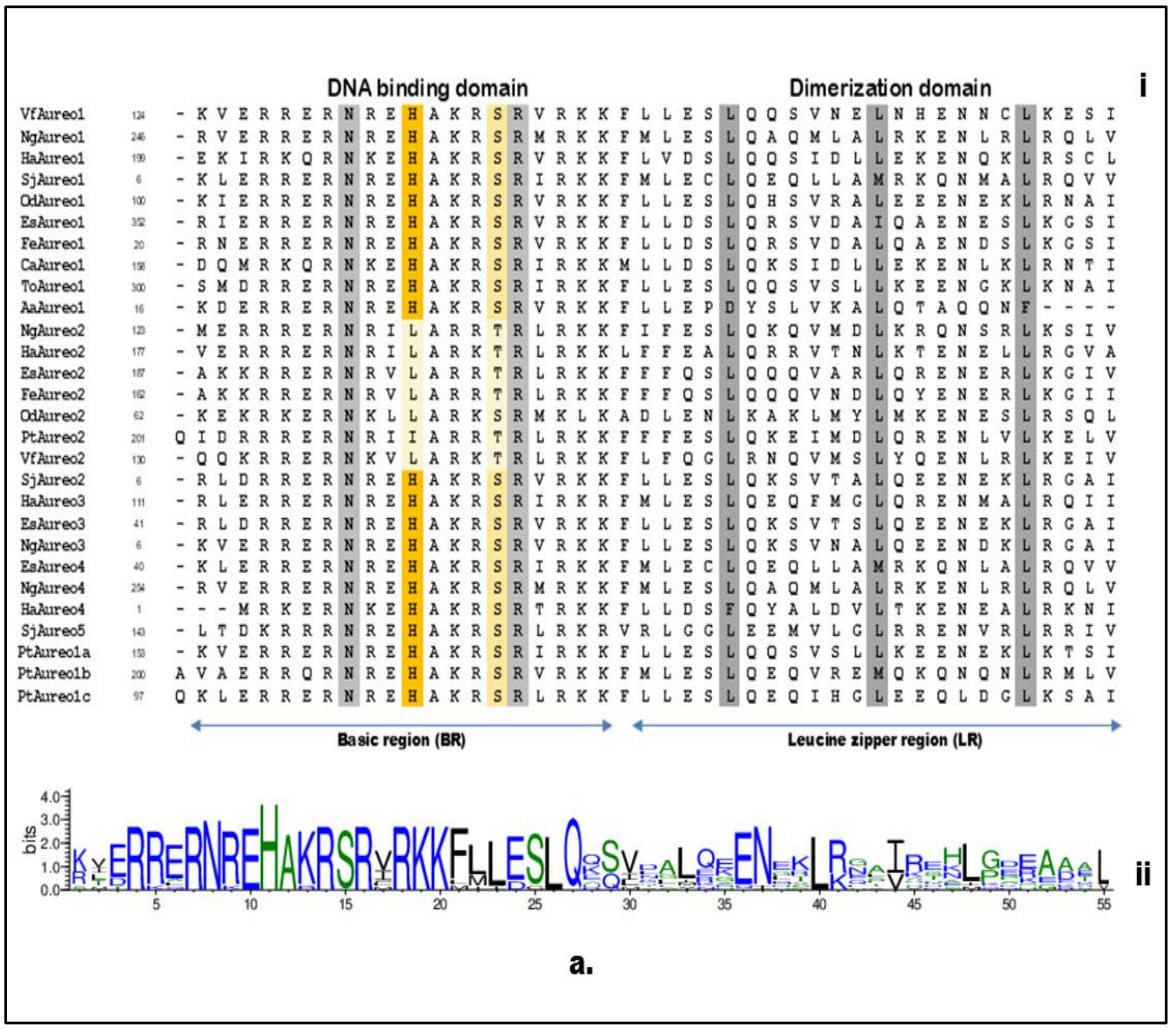
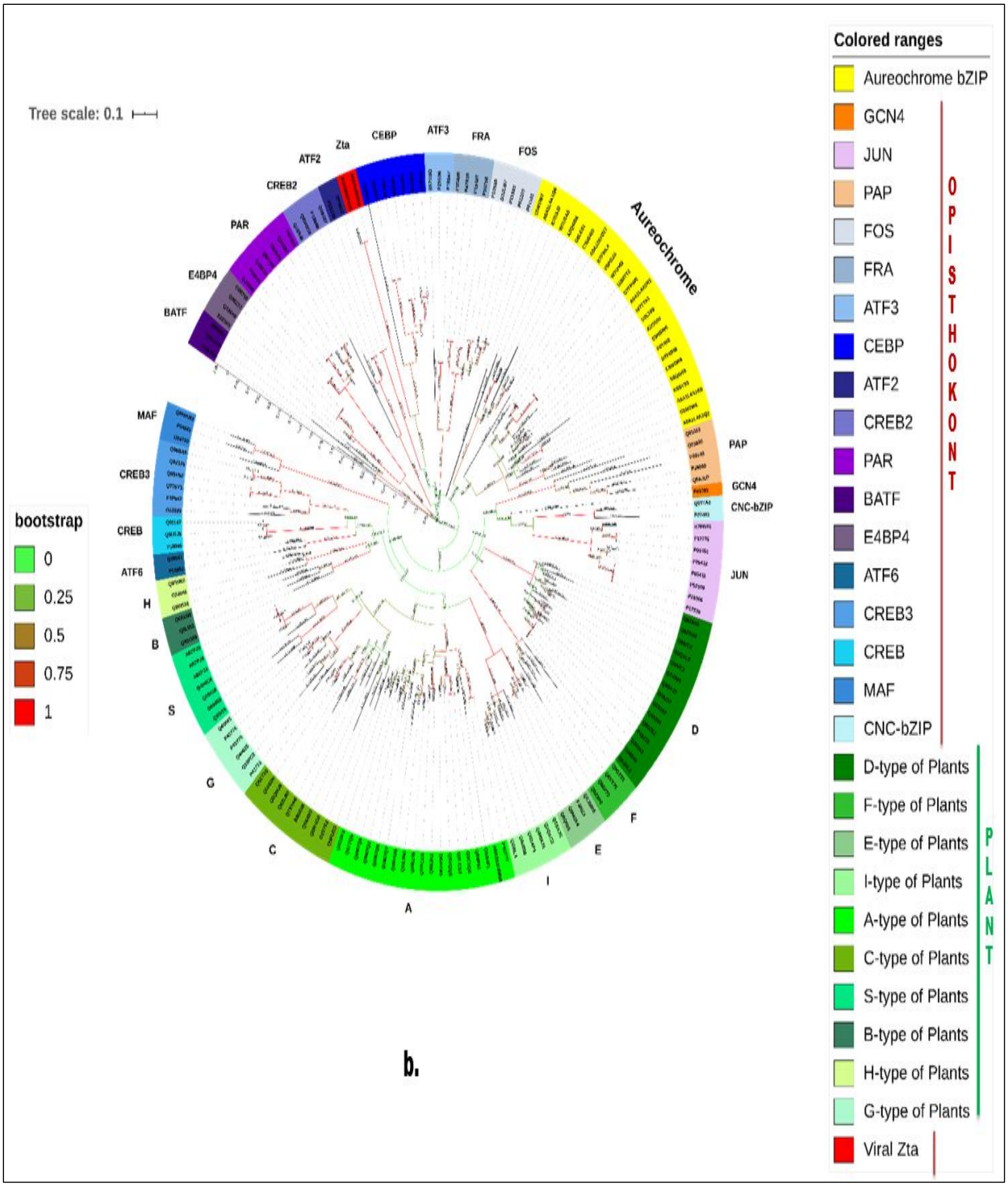


Figure-1: a. Phylogenetic tree b. Domain architecture of all aureochromes. Representation of the phylogeny of the domains of aureochromes separately c. bZIP d. LOV.

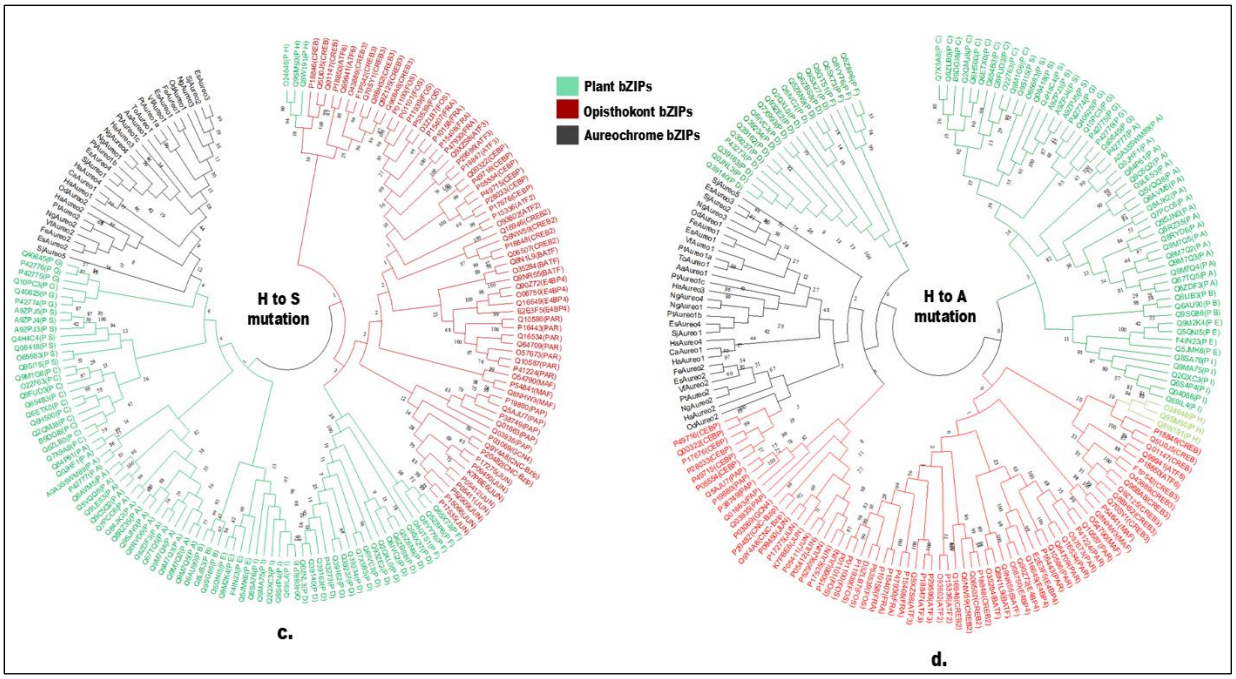
MSA of bZIP domains pointed out the differences in residue composition among the different subtypes of aureo-bZIPs [**Figure-2.a.i**]. Since aureochrome represented the only light responsive bZIP, sequences of all other eukaryotic bZIPs were compared with it to determine its differences with conventional bZIPs. Though the conservation of asparagine and arginine as the 1st and 9th (last) residue of the signature motif of DNA binding region was found, 4th residue of aureo-bZIPs comprised histidine (except all aureo2) unlike other bZIPs. Co-evolution or co-variance studies of aureo-bZIP keeping an E-value cut-off 10^{-10} indicatesd distinct presence of histidine [**Figure-2.a.ii**]. E-value when increased to 10^{-4} for accommodating more bZIP sequences, complete disappearance of histidine was observed. The phylogenetic tree of all bZIPs [**Figure-2.b**] reflectsed the tendency of aureochrome-bZIPs to club together forming a distinct clade. Histidine, when mutated to serine (as found in plants) and alanine (as found in opisthokont) aureochrome clade shifted closer to plant

and opisthokonts bZIPs respectively [**Figure-2.c, 2.d**]. In fact, the residue present in the corresponding positions of plants/opisthokonts upon substitution to histidine resulted in positioning of aureochrome-bZIPs under a common ancestral lineage with plants and opisthokont-bZIPs respectively [**Figure-2.e, 2.f**].


DNA binding domain
Dimerization domain
i
VfAureo1 124 - K V E R R E R N R E H A K R S R V R K K F L L E S L Q Q S V N E L N H E N N C L K E S I
NgAureo1 246 - R V E R R E R N R E H A K R S R M R K K F M L E S L Q A Q M L A L R K E N L R L R Q L V
HaAureo1 199 - E K I R K Q R N K E H A K R S R V R K K F L V D S L Q Q S I D L L E K E N Q K L R S C L
SjAureo1 6 - K L E R R E R N R E H A K R S R I R K K F M L E C L Q E Q L L A M R K Q N M A L R Q V V
OdAureo1 100 - K I E R R E R N R E H A K R S R V R K K F L L E S L Q H S V R A L E E E N E K L R N A I
EsAureo1 352 - R I E R R E R N R E H A K R S R V R K K F L L D S L Q R S V D A I Q A E N E S L K G S I
FeAureo1 20 - R N E R R E R N R E H A K R S R V R K K F L L D S L Q R S V D A L Q A E N D S L K G S I
CaAureo1 198 - D Q M R K Q R N K E H A K R S R I R K K M L L D S L Q K S I D L L E K E N L K L R N T I
ToAureo1 300 - S M D R R E R N R E H A K R S R I R K K F L L E S L Q Q S V S L L K E E N G K L K N A I
AaAureo1 16 - K D E R R E R N R E H A K R S R V R K K F L L E P D Y S L V K A L Q T A Q Q N F - - - -
NgAureo2 123 - M E R R R E R N R I L A R R T R L R K K F I F E S L Q K Q V M D L K R Q N S R L K S I V
HaAureo2 177 - V E R R R E R N R I L A R K T R L R K K L F F E A L Q R R V T N L K T E N E L L R G V A
EsAureo2 187 - A K K R R E R N R V L A R R T R L R K K F F F Q S L Q Q Q V A R L Q R E N E R L K G I V
FeAureo2 162 - A K K R R E R N R V L A R R T R L R K K F F F Q S L Q Q Q V N D L Q Y E N E R L K G I I
OdAureo2 62 - K E K R K E R N K L L A R K S R M K L K A D L E N L K A K L M Y L M K E N E S L R S Q L
PtAureo2 201 Q I D R R R E R N R I I A R R T R L R K K F F F E S L Q K E I M D L Q R E N L V L K E L V
VfAureo2 130 - Q Q K R R E R N K V L A R K T R L R K K F L F Q G L R N Q V M S L Y Q E N L R L K E I V
SjAureo2 6 - R L D R R E R N R E H A K R S R V R K K F L L E S L Q K S V T A L Q E E N E K L R G A I
HaAureo3 111 - R L E R R E R N R E H A K R S R I R K R F M L E S L Q E Q F M G L Q R E N M A L R Q I I
EsAureo3 41 - R L D R R E R N R E H A K R S R V R K K F L L E S L Q K S V T S L Q E E N E K L R G A I
NgAureo3 6 - K V E R R E R N R E H A K R S R V R K K F L L E S L Q K S V N A L Q E E N D K L R G A I
EsAureo4 40 - K L E R R E R N R E H A K R S R I R K K F M L E C L Q E Q L L A M R K Q N L A L R Q V V
NgAureo4 254 - R V E R R E R N R E H A K R S R M R K K F M L E S L Q A Q M L A L R K E N L R L R Q L V
HaAureo4 1 - - - M R K E R N K E H A K R S R T R K K F L L D S F Q Y A L D V L T K E N E A L R K N I
SjAureo5 143 - L T D K R R R N R E H A K R S R L R K R V R L G G L E E M V L G L R R E N V R L R R I V
PtAureo1a 153 - K V E R R E R N R E H A K R S R I R K K F L L E S L Q Q S V S L L K E E N E K L K T S I
PtAureo1b 200 A V A E R R Q R N R E H A K R S R V R K K F M L E S L Q E Q V R E M Q K Q N Q N L R M L V
PtAureo1c 97 Q K L E R R E R N R E H A K R S R L R K K F L L E S L Q E Q I H G L E E Q L D G L K S A I
Basic region (BR)
Leucine zipper region (LR)
ii
bits
4.0
3.0
2.0
1.0
0.0
5
10
15
20
25
30
35
40
45
50
55
a.


Tree scale: 0.1
bootstrap
0
0.25
0.5
0.75
1
Colored ranges
Aureochrome bZIP
GCN4
JUN
PAP
FOS
FRA
ATF3
CEBP
ATF2
CREB2
PAR
BATF
E4BP4
ATF6
CREB3
CREB
MAF
CNC-bZIP
OPISTHOKONT
D-type of Plants
F-type of Plants
E-type of Plants
I-type of Plants
A-type of Plants
C-type of Plants
S-type of Plants
B-type of Plants
H-type of Plants
G-type of Plants
PLANT
Viral Zta
Aureochrome
PAP
GCN4
CNC-bZIP
JUN
D
F
E
I
A
C
G
S
B
H
ATF6
CREB
CREB3
MAF
BATF
E4BP4
PAR
CREB2
ATF2
Zta
CEBP
ATF3
FRA
FOS

b.



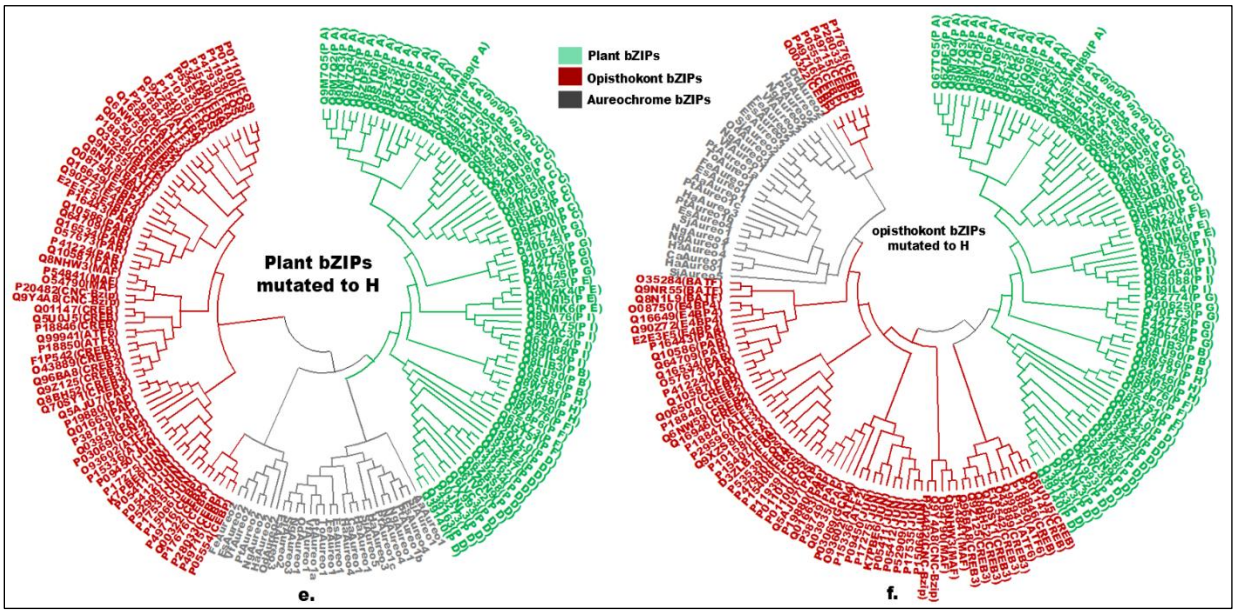


Figure-2: a. i- Multiple sequence alignment of the bZIP domains of Aureos. Conserved residues (Asn and Arg) that outline the signature basic region as well as the 'd' positions (mostly Leu) of the heptads are highlighted in gray. The 4th and 8th residues of the basic regions are histidine and serine respectively in all Aureo homologs, except for Aureo2 orthologs. ii- Weblogo result of the co-evolution study of aureochrome1 at E-value cut-off 10^{-10} depicting the presence of unique histidine at the basic region.

b. Phylogenetic tree constructed with the bZIPs from plants, opisthokonts and Aureos. Different subgroups of bZIPs are represented by different colours as indicated. The branches are color coded from green to red in

the order of highest to lowest bootstrap values. Aureochrome homologs (highlighted in yellow) are found on a clade separate from that of the plants and opisthokonts.

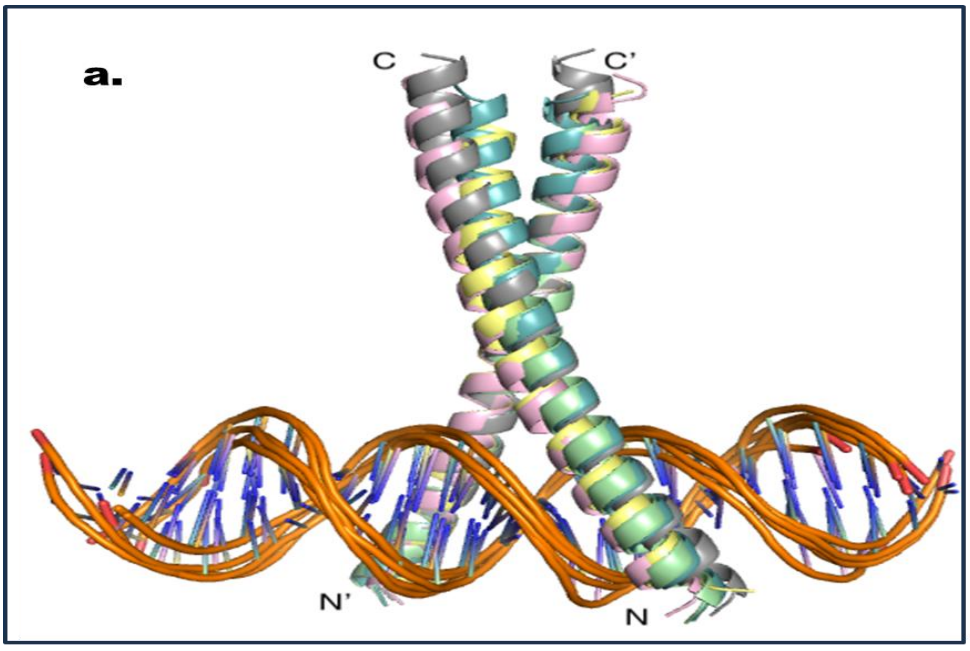
c. Re-construction of phylogenetic tree upon mutation of Aureos' histidine to serine (plant like). Serine substituted Aureo bZIPs are found to club with plant bZIPs (barring plant bZIP d, f and h) in the re-constructed phylogenetic tree.

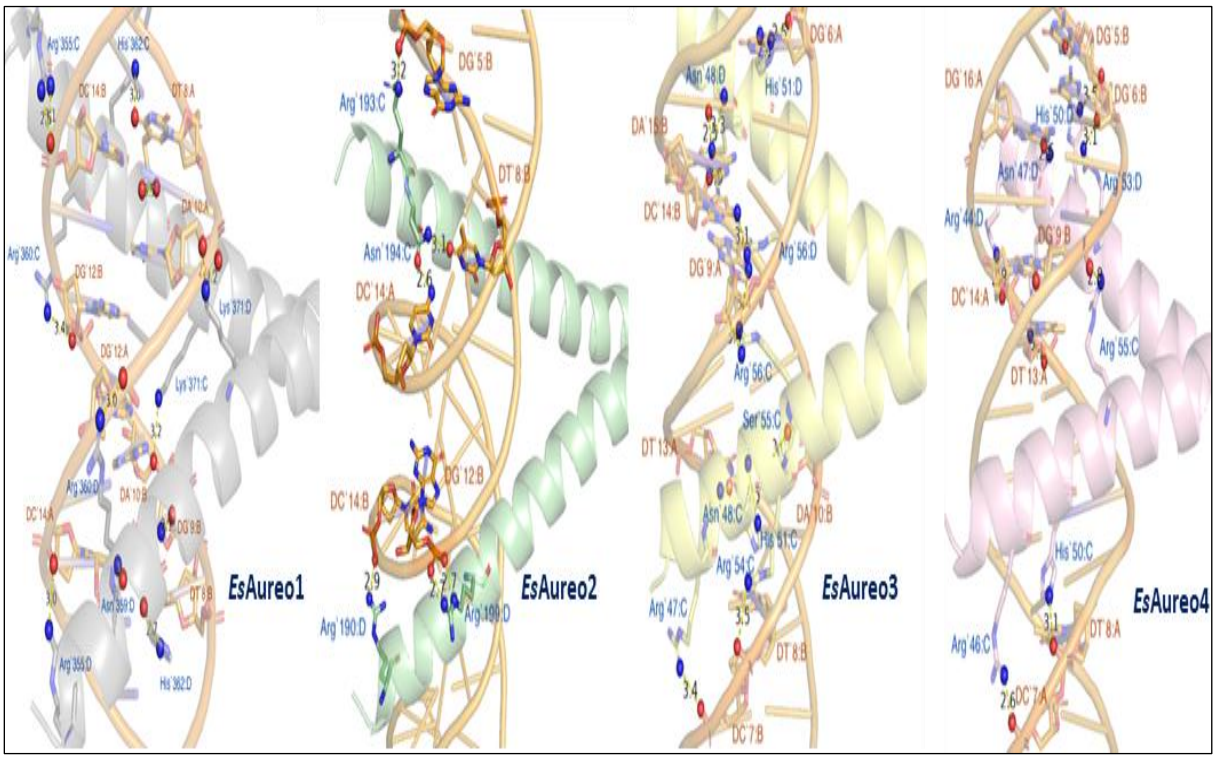
d. Re-construction of phylogenetic tree upon mutation of Aureos' histidine to alanine (opisthokont like). Alanine substituted Aureo bZIPs are seen to occupy a region nearby opisthokont bZIPs (except CEBP, JUN, PAP and CNC-bZIP from opisthokonts) in the reconstructed phylogenetic tree.

e. Re-construction of the phylogenetic tree by incorporating histidine in place of serine in plant bZIPs. The plant bZIPs are seen to re-group with the Aureo bZIPs under a common ancestral lineage upon single histidine substitution.

f. Re-construction of the phylogenetic tree by incorporating histidine in place of alanine in opisthokont bZIPs. The opisthokont bZIPs (except CEBP bZIP sub-family) are found to re-group with the Aureo bZIPs under a common ancestral lineage.

The modelled bZIPs docked with CRE-DNA exhibited polar interaction of histidine with the substrate DNA except in aureo2 [**Figure-3.b**]. Leucine present in aureo-2 failed to interact with DNA.





b

Figure-3: a. Modelled bZIP structures of *Ectocarpus siliculosus* (*Es*) Aureo1 (gray), Aureo2 (pale green), Aureo3 (yellow), Aureo4 (pink) and Saccharina japonica (*Sj*) Aureo5 (teal) bZIPs docked with the substrate DNA sequences containing 'Aureo-box', TGACGT motif. b. Polar interactions between amino acid residues from the basic region of bZIPs and the corresponding nucleotide from the substrate DNA are labeled. Interacting residues from protein as well as DNA are represented in sticks whereas the nitrogen and oxygen atoms participating in non-covalent interactions, predominantly hydrogen bonding/polar interactions, are marked with blue and red spheres respectively. Non-covalent interactions within a distance cut off of 3.5 Å are displayed.

No significant interactions were found if histidine is mutated to serine/alanine. Dynamics of hydrogen bonding profile obtained through MD simulation represents greater number of hydrogen bonds with DNA throughout the 60ns time scale and confirmed the importance of histidine over alanine/serine for maintaining stable interaction with the DNA substrate [**Figure-4**].

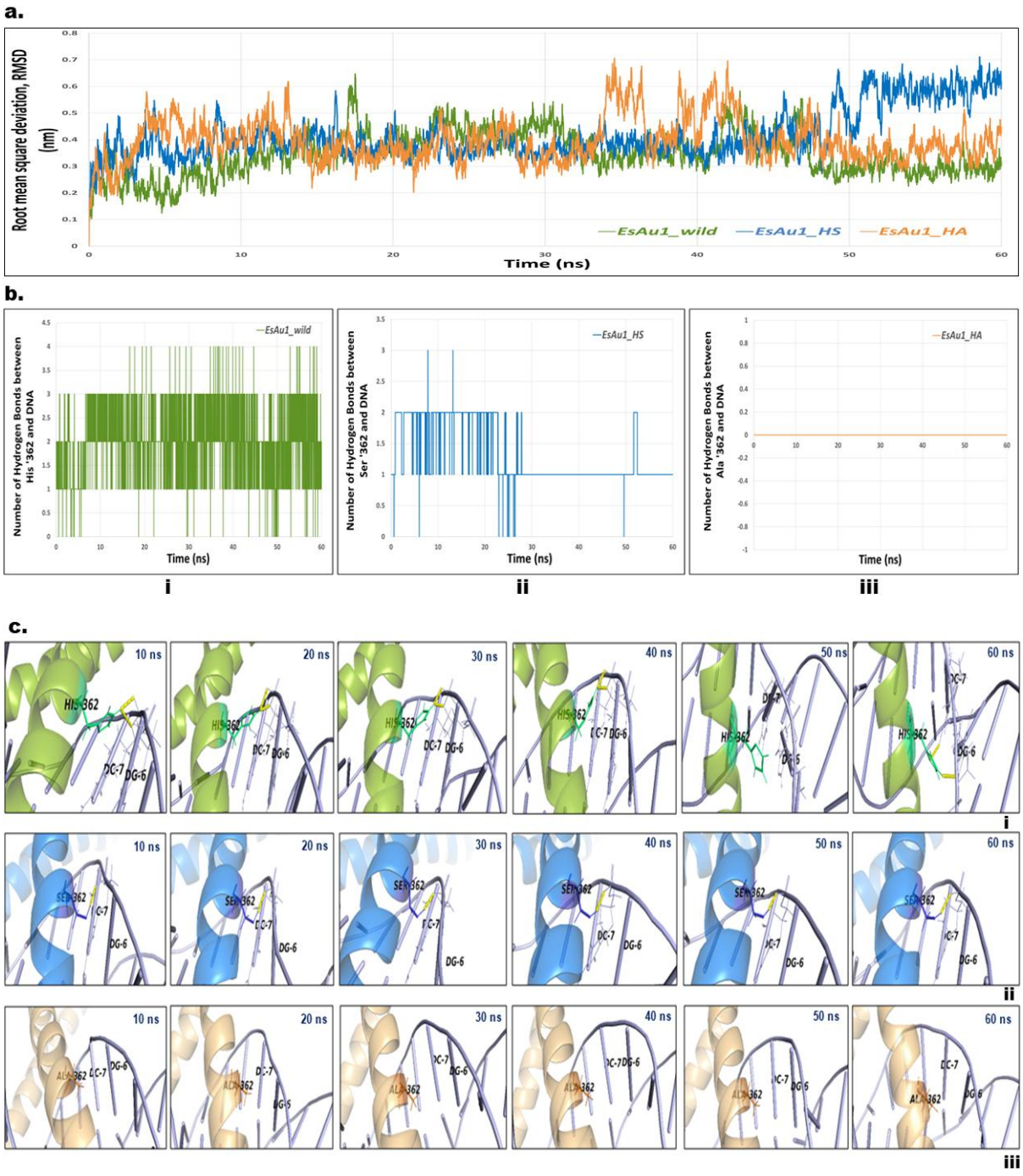


Figure-4: a. Root Mean Square Deviation (RMSD) in nm plotted for EsAureo1_wild (green), EsAureo1_HS (blue) and EsAureo1_HA (orange) protein backbone for 60 ns timescale of MD simulation. b. Average number of intermolecular hydrogen bonds during 60 ns simulation timescale. i. between His362 in EsAureo1_wild and DNA. ii. between Ser362 in EsAureo1_HS and DNA. iii. between Ala362 in EsAureo1_HA and DNA. c. Post-simulation visualization of residue 362 (histidine) and its hydrogen bonding interaction with DNA (shown in yellow dashed line), with frames taken at 10 ns intervals (from left to right).

i.EsAureo1_wild (green), where His362 maintains mostly consistent interaction with the DNA substrate (nucleotides highlighted). ii.EsAureo1_HS (blue), with Ser362 showing a yet consistent but weaker contact with the substrate DNA (nucleotides highlighted). iii.EsAureo1_HA (orange), with Ala362 although being present in close proximity to the nitrogenous bases of DNA (nucleotides highlighted), fails to establish any contact.

In-vitro DNA binding study by EMSA [**Figure-5]** also exhibited increased DNA binding affinity of histidine containing wild-type protein with an average K_D of 0.13 ± 0.03 µM than the mutants (K_D - 0.24 ± 0.08 µM and 1.1 ± 0.87 µM).

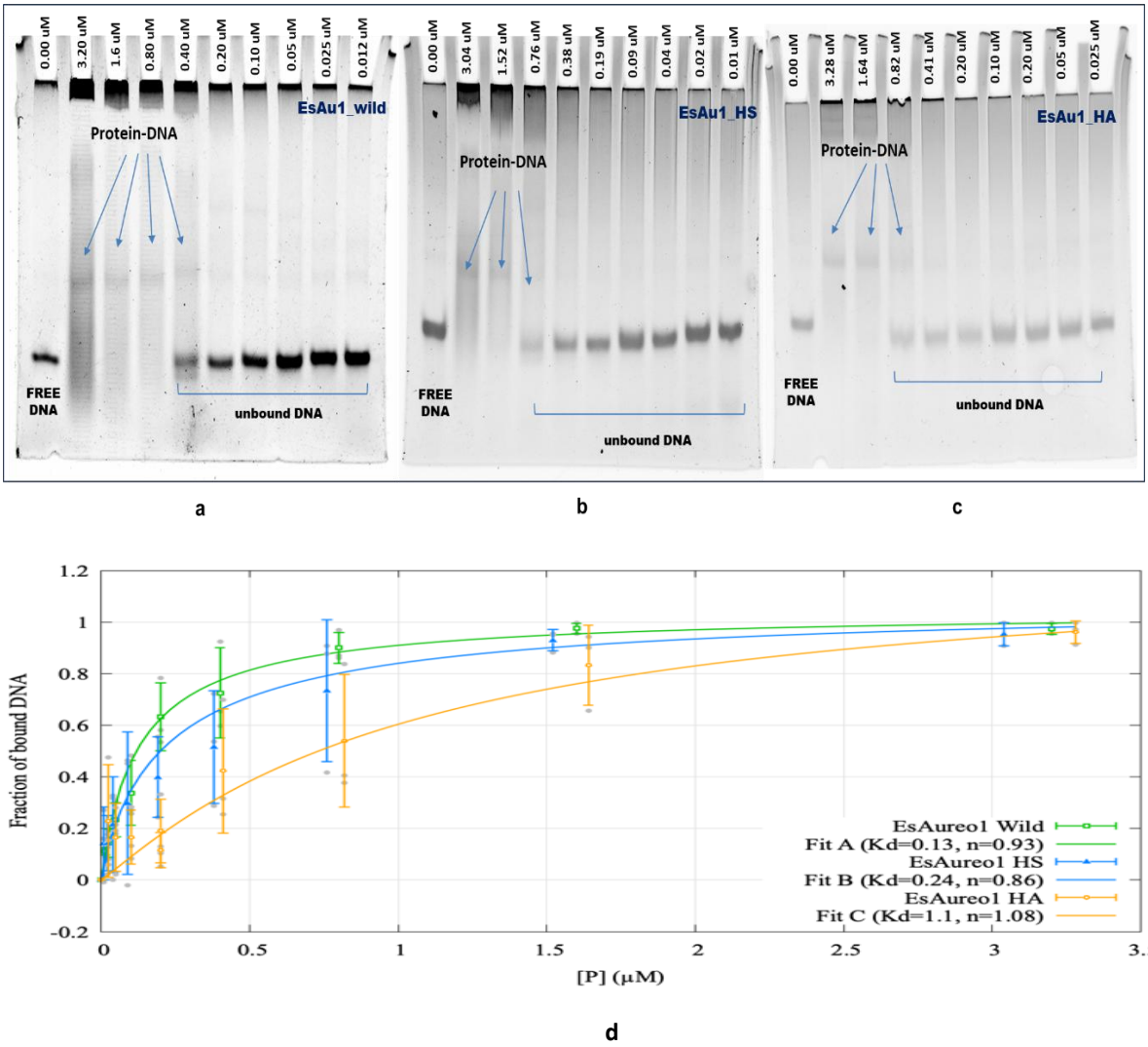
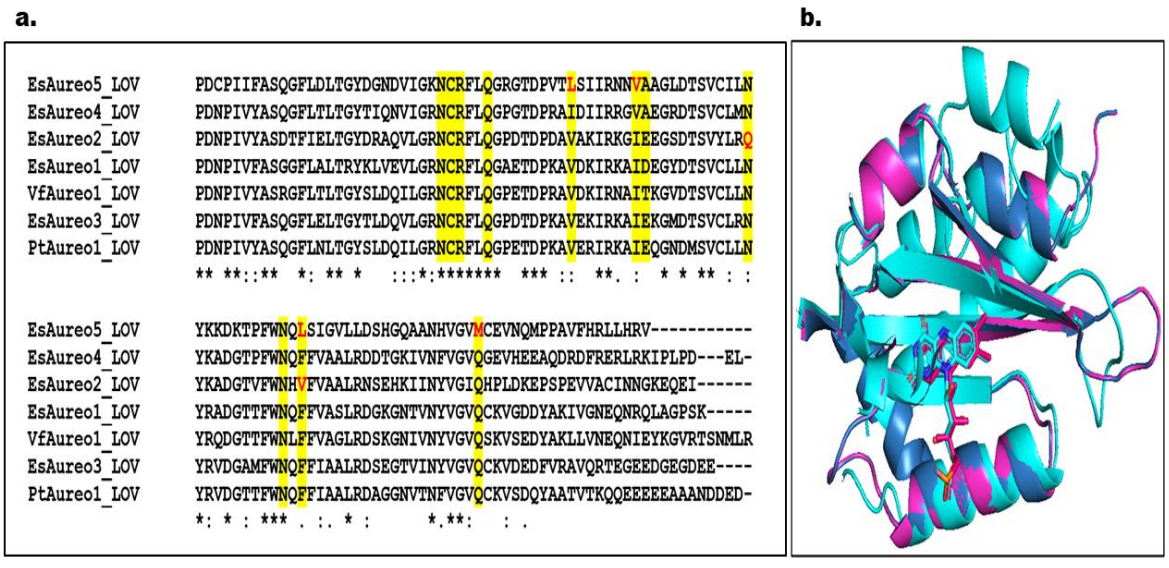


Figure-5: In vitro protein-DNA interaction study of Aureochrome bZIPs through electrophoretic mobility shift assay-a. *Es*Aureo1_wild, b. *Es*Aureo_HS, c. *Es*Aureo_HA. d. DNA-binding affinity (KD) for *Es*Aureo1_wild, *Es*Aureo_HS and *Es*Aureo_HA determined from ImageJ analysis followed by Hill curve

fitting. For the plotted curves, grey dots represent the actual value while the green, blue and orange markers with whiskers denote the average fraction of bound DNA values of EsAureo1_wild, EsAureo_HS and *Es*Aureo_HA respectively (continuous lines representing fitted data) against different protein concentrations (in μM). KD values obtained are 0.13 ± 0.03 μM, 0.24 ± 0.08 μM and 1.1 ± 0.87 μM respectively for *Es*Aureo1_wild, *Es*Aureo_HS and *Es*Aureo_HA. The wild-type protein shows a marked DNA-binding affinity while the mutants have a comparatively lesser binding strength, thereby highlighting the role of histidine in target DNA recognition and specific-binding profile of EsAureo1 protein.

A thorough scrutiny into the LOV sensor sequences of all the five types of aureochromes found in *E. siliculosus* using *in-silico* tools from structural bioinformatics revealed distinct variations in the LOV domain sequences of aureochrome-5 and aureochrome-2 in the conserved FMN binding region [**Figure-6.a**]. LOV domains of aureochrome-1, 3 and 4 exhibited similarities in residue composition. Superimposed models of LOV from *Es*Aureo1, *Es*Aureo2 and *Es*Aureo5 with RMS value 0.99 and 1.01 for *Es*Aureo1-*Es*Aureo2 and *Es*Aureo1-*Es*Aureo5 respectively [**Figure-6.b**] indicated structural differences. Analysis of the hydrogen bonding and hydrophobic interactions of the proteins with the incorporated FMN [**Figure-6.c**] as a representation of the FMN binding environments of these aureochrome paralogs depicted significant hydrogen bonding and hydrophobic interactions of *Es*Aureo1 protein with FMN that indicated stable FMN binding of *Es*Aureo1 protein than the rest.



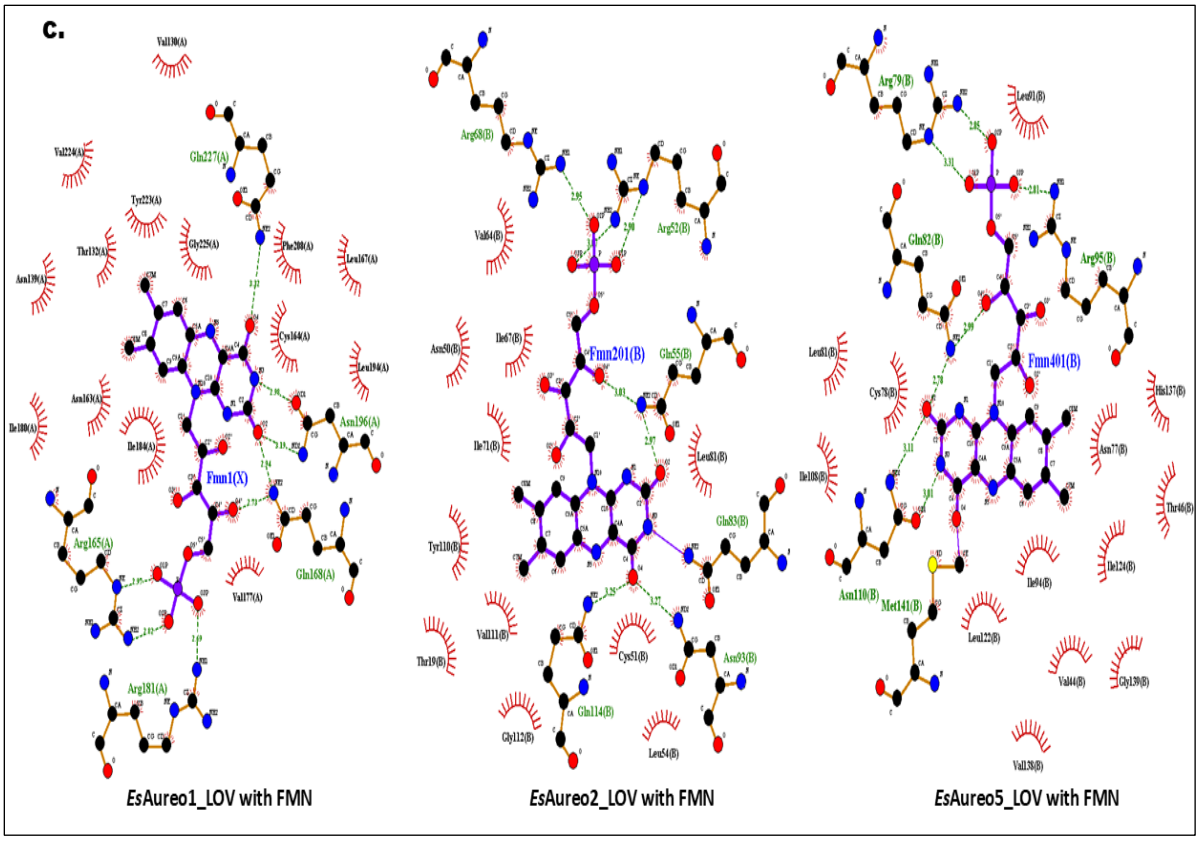


Figure-6: a. Multiple sequence alignment of the LOV domains of aureochromes. b. superimposed models of *Es*Aureo1, *Es*Aureo2 and *Es*Aureo5 proteins c. FMN binding environment of *Es*Aureo1, *Es*Aureo2 and *Es*Aureo5 proteins

An in-depth *in-vitro* study of *Es*Aureo1 as the representative of a typical aureochrome family and *Es*Aureo2, *Es*Aureo5 as the strikingly different variant attempted structural and functional characterization. Long disordered N-terminal region was observed upstream of bZIP upon modeling of *Es*Aureo1 protein in alpha fold 3. So, only the codon optimized 687bp long synthetic gene corresponding to the bZIP and LOV domain along with the linker of *Es*Aureo1 was cloned into pET28a expression vector. The purified overexpressed *Es*Aureo1 showed a triplet absorption peak around 447nm at dark during UV-Vis spectroscopy [**Figure-7.a**] that disappeared after 12 mins of illumination. The presence of protein bands in SDS PAGE at a position corresponding to its molecular weight confirmed the homogeneity of the protein [**Figure-7.b**].

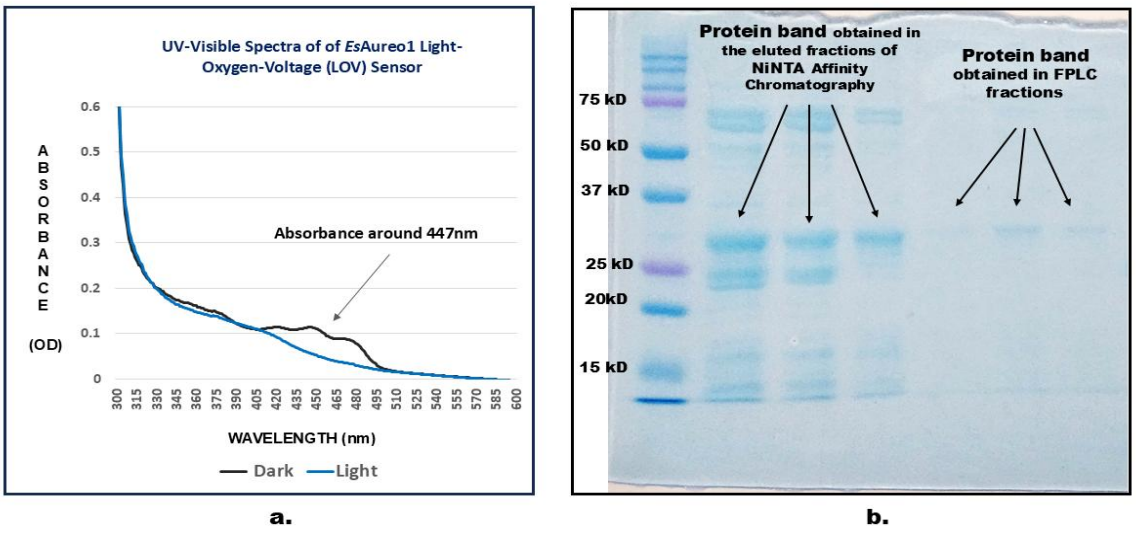
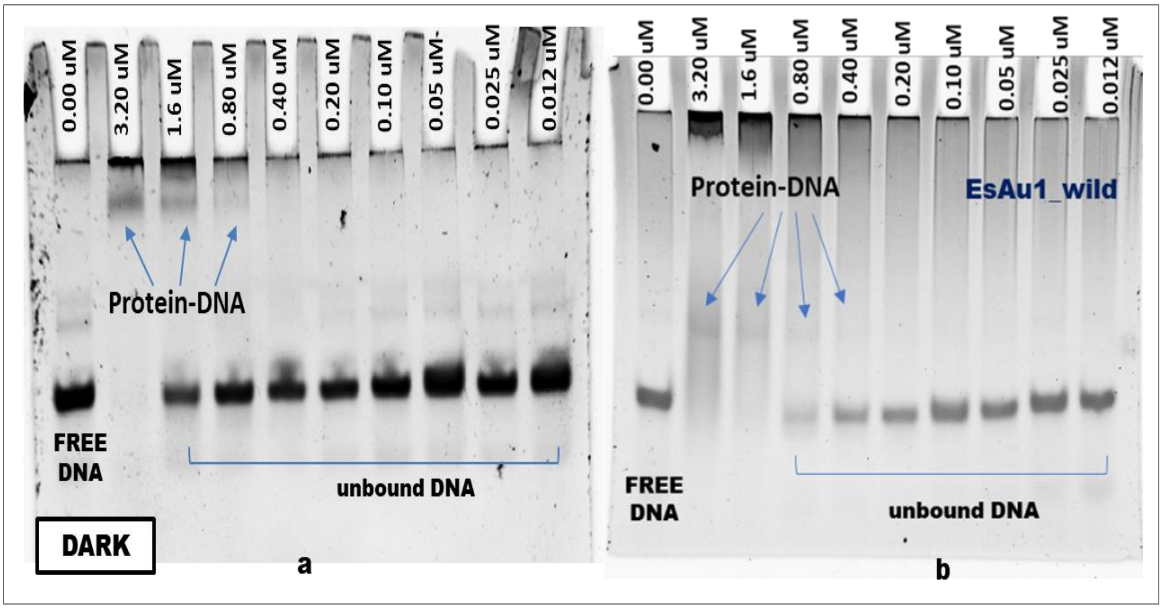


Figure-7: a. UV-Visible absorption spectra of a functional LOV sensor (dark versus photobleached or states) of EsAureo1 protein. The black line depicts the spectrum of dark state with a typical absorption maximum around 447nm. The blue line represents the photobleached state obtained after 10 mins of continuous blue light exposure. b. SDS gel image showing bands for *E*sAureo1 protein

Efficient DNA binding in presence and absence of light in EMSA with difference in affinity proved its functionality [**Figure-8**].



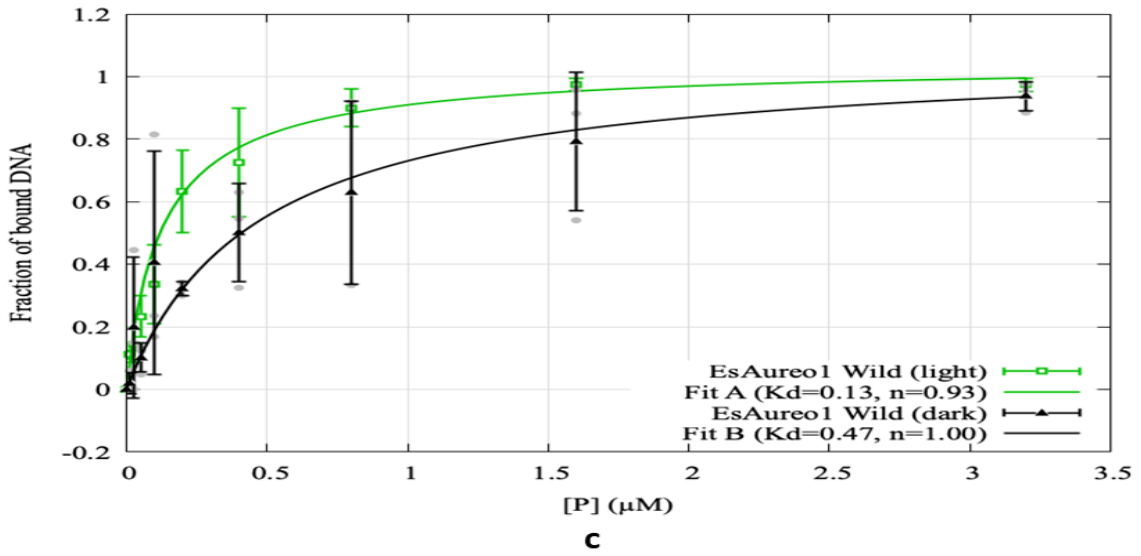
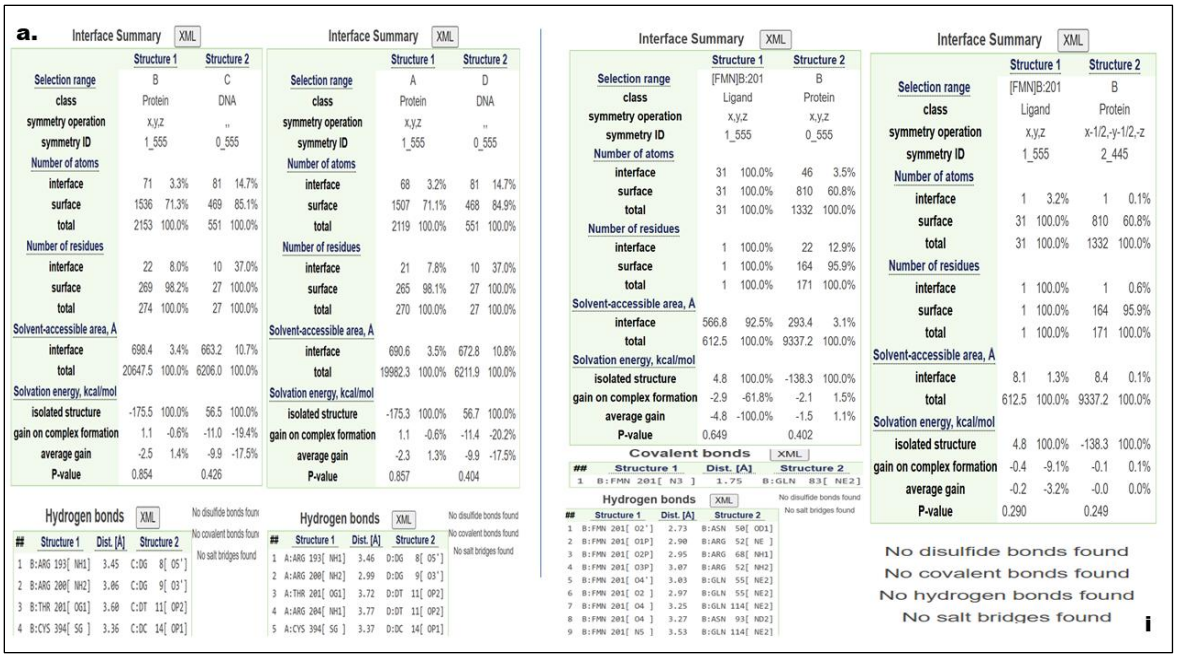
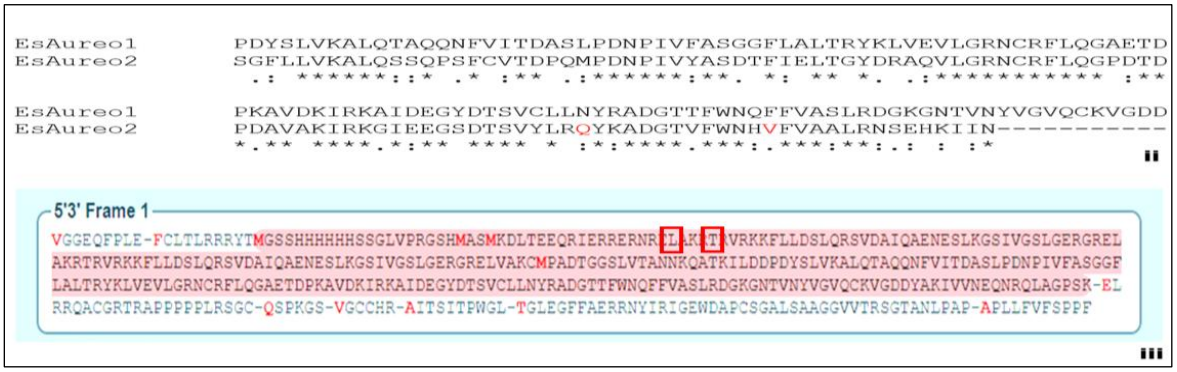


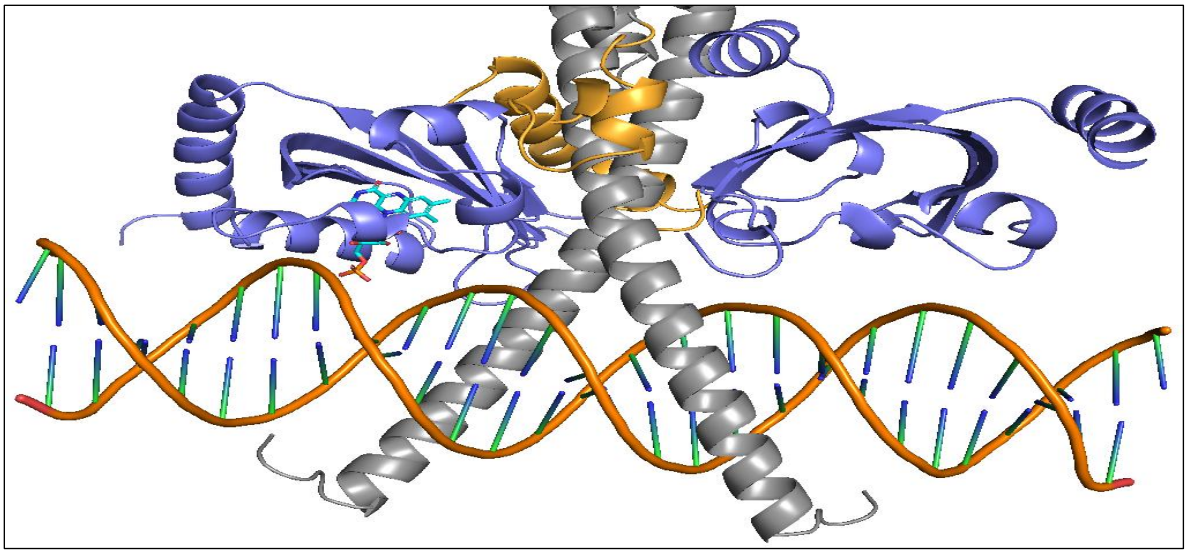
Figure-8: EMSA gel images of EsAureo1_Wild in (i) dark state and (ii) light state, with (iii) K_D value estimation and comparison of light versus dark states. Green and black markers accompanied with whiskers denote the average fraction of bound DNA values of *Es*Aureo1_wild protein at light and dark respectively (continuous lines representing fitted data) against different protein concentrations (in μM). K_D values of 0.13 ± 0.03 μM and 0.47 ± 0.08 μM respectively for light and dark states indicate enhanced DNA-binding affinity of Aureo1_Wild under blue light conditions.

In silico tools of structural bioinformatics identified interactions of *Es*Aureo2 protein models with DNA [**Figure-9.a.i**] like *Es*Aureo1. No covalent, disulfide or hydrogen bonding interactions between the modeled protein *Es*Aureo2 and FMN was found unlike *Es*Aureo1. This finding corroboratesd well with the literature that stated aureochrome-2 homologs do not form stable binding with FMN and was considered as non-functional. MSA revealed differences in protein sequences of *Es*Aureo1 and *Es*Aureo2 with crucial FMN binding residue substitution in *Es*Aureo2 [**Figure-9.a.i**]. Striking residue substitution in the DNA binding motif was also observed. *Es*Aureo2 like DNA obtained through the incorporation of mutations in *Es*Aureo1 by site directed mutagenesis [**Figure-9.a.ii**] was subjected to protein over-expression, purification [**Figure-9.b.ii-iii**] and EMSA [**Figure-9.c.(i-iii)**] to check the functionality of its bZIP domain. *Es*Aureo2 like protein exhibited DNA binding with less affinity than *Es*Aureo1. *Es*Aureo2 protein was stable and formed hydrogen bonding interaction with DNA was validated by MD simulation data [**Figure-9.d(i-iv)**].

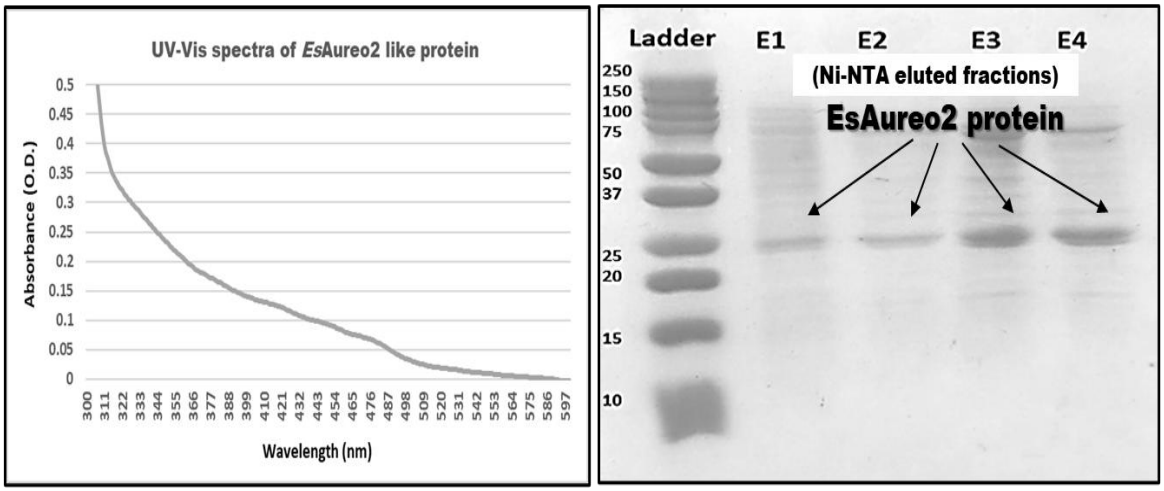

a.
Interface Summary
XML
Structure 1
Structure 2
Selection range
B
C
class
Protein
DNA
symmetry operation
x,y,z
,,
symmetry ID
1_555
0_555
Number of atoms
interface 71 3.3% 81 14.7%
surface 1536 71.3% 469 85.1%
total 2153 100.0% 551 100.0%
Number of residues
interface 22 8.0% 10 37.0%
surface 269 98.2% 27 100.0%
total 274 100.0% 27 100.0%
Solvent-accessible area, A
interface 698.4 3.4% 663.2 10.7%
total 20647.5 100.0% 6206.0 100.0%
Solvation energy, kcal/mol
isolated structure -175.5 100.0% 56.5 100.0%
gain on complex formation 1.1 -0.6% -11.0 -19.4%
average gain -2.5 1.4% -9.9 -17.5%
P-value 0.854 0.426
Hydrogen bonds
XML
No disulfide bonds found
No covalent bonds found
No salt bridges found
Structure 1 Dist. [A] Structure 2
1 B:ARG 193[NH1] 3.45 C:DG 8[O5']
2 B:ARG 200[NH2] 3.06 C:DG 9[O3']
3 B:THR 201[OG1] 3.60 C:DT 11[OP2]
4 B:CYS 394[SG] 3.36 C:DC 14[OP1]
Interface Summary
XML
Structure 1
Structure 2
Selection range
A
D
class
Protein
DNA
symmetry operation
x,y,z
,,
symmetry ID
1_555
0_555
Number of atoms
interface 68 3.2% 81 14.7%
surface 1507 71.1% 468 84.9%
total 2119 100.0% 551 100.0%
Number of residues
interface 21 7.8% 10 37.0%
surface 265 98.1% 27 100.0%
total 270 100.0% 27 100.0%
Solvent-accessible area, A
interface 690.6 3.5% 672.8 10.8%
total 19982.3 100.0% 6211.9 100.0%
Solvation energy, kcal/mol
isolated structure -175.3 100.0% 56.7 100.0%
gain on complex formation 1.1 -0.6% -11.4 -20.2%
average gain -2.3 1.3% -9.9 -17.5%
P-value 0.857 0.404
Hydrogen bonds
XML
No disulfide bonds found
No covalent bonds found
No salt bridges found
Structure 1 Dist. [A] Structure 2
1 A:ARG 193[NH1] 3.46 D:DG 8[O5']
2 A:ARG 200[NH2] 2.99 D:DG 9[O3']
3 A:THR 201[OG1] 3.72 D:DT 11[OP2]
4 A:ARG 204[NH1] 3.77 D:DT 11[OP2]
5 A:CYS 394[SG] 3.37 D:DC 14[OP1]
Interface Summary
XML
Structure 1
Structure 2
Selection range
[FMN]B:201
B
class
Ligand
Protein
symmetry operation
x,y,z
x,y,z
symmetry ID
1_555
0_555
Number of atoms
interface 31 100.0% 46 3.5%
surface 31 100.0% 810 60.8%
total 31 100.0% 1332 100.0%
Number of residues
interface 1 100.0% 22 12.9%
surface 1 100.0% 164 95.9%
total 1 100.0% 171 100.0%
Solvent-accessible area, A
interface 566.8 92.5% 293.4 3.1%
total 612.5 100.0% 9337.2 100.0%
Solvation energy, kcal/mol
isolated structure 4.8 100.0% -138.3 100.0%
gain on complex formation -2.9 -61.8% -2.1 1.5%
average gain -4.8 -100.0% -1.5 1.1%
P-value 0.649 0.402
Covalent bonds
XML
Structure 1 Dist. [A] Structure 2
1 B:FMN 201[N3] 1.75 B:GLN 83[NE2]
Hydrogen bonds
XML
No disulfide bonds found
No salt bridges found
Structure 1 Dist. [A] Structure 2
1 B:FMN 201[O2'] 2.73 B:ASN 50[OD1]
2 B:FMN 201[O1P] 2.90 B:ARG 52[NE]
3 B:FMN 201[O2P] 2.95 B:ARG 68[NH1]
4 B:FMN 201[O3P] 3.07 B:ARG 52[NH2]
5 B:FMN 201[O4'] 3.03 B:GLN 55[NE2]
6 B:FMN 201[O2] 2.97 B:GLN 55[NE2]
7 B:FMN 201[O4] 3.25 B:GLN 114[NE2]
8 B:FMN 201[O4] 3.27 B:ASN 93[ND2]
9 B:FMN 201[N5] 3.53 B:GLN 114[NE2]
Interface Summary
XML
Structure 1
Structure 2
Selection range
[FMN]B:201
B
class
Ligand
Protein
symmetry operation
x,y,z
x-1/2,-y-1/2,-z
symmetry ID
1_555
2_445
Number of atoms
interface 1 3.2% 1 0.1%
surface 31 100.0% 810 60.8%
total 31 100.0% 1332 100.0%
Number of residues
interface 1 100.0% 1 0.6%
surface 1 100.0% 164 95.9%
total 1 100.0% 171 100.0%
Solvent-accessible area, A
interface 8.1 1.3% 8.4 0.1%
total 612.5 100.0% 9337.2 100.0%
Solvation energy, kcal/mol
isolated structure 4.8 100.0% -138.3 100.0%
gain on complex formation -0.4 -9.1% -0.1 0.1%
average gain -0.2 -3.2% -0.0 0.0%
P-value 0.290 0.249
No disulfide bonds found
No covalent bonds found
No hydrogen bonds found
No salt bridges found
i


EsAureo1 PDYSLVKALQTAQQNFVITDASLPDNPIVFASGGFLALTRYKLVEVLGRNCRFLQGAETD
EsAureo2 SGFLLVKALQSSQPSFCVTDPQMPDNPIVYASDTFIELTGYDRAQVLGRNCRFLQGPDTD
EsAureo1 PKAVDKIRKAIDEGYDTSVCLLNYRADGTTFWNQFFVASLRDGKGNTVNYVGVQCKVGDD
EsAureo2 PDAVAKIRKGIEEGSDTSVYLRQYKADGTVFWNHVFVAALRNSEHKIIN----------
5'3' Frame 1
VGGEQFPLE-FCLTLRRRYTMGSSHHHHHHSSGLVPRGSHMASMKDLTEEQRIERRERNRLLKRTRVRKKFLLDSLQRSVDAIQAENESLKGSIVGSLGERGREL
AKRTRVRKKFLLDSLQRSVDAIQAENESLKGSIVGSLGERGRELVAKCMPADTGGSLVTANNKQATKILDDPDYSLVKALQTAQQNFVITDASLPDNPIVFASGGF
LALTRYKLVEVLGRNCRFLQGAETDPKAVDKIRKAIDEGYDTSVCLLNYRADGTTFWNQFFVASLRDGKGNTVNYVGVQCKVGDDYAKIVVNEQNRQLAGPSK-EL
RRQACGRTRAPPPPPLRSGC-QSPKGS-VGCCHR-AITSITPWGL-TGLEGFFAERRNYIRIGEWDAPCSGALSAAGGVVTRSGTANLPAP-APLLFVFSPPF

a

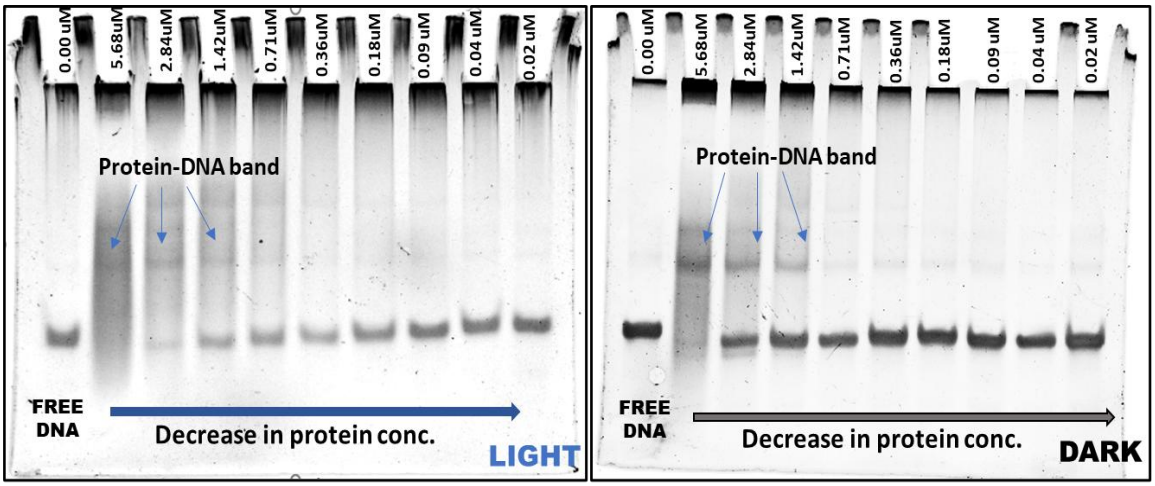


b

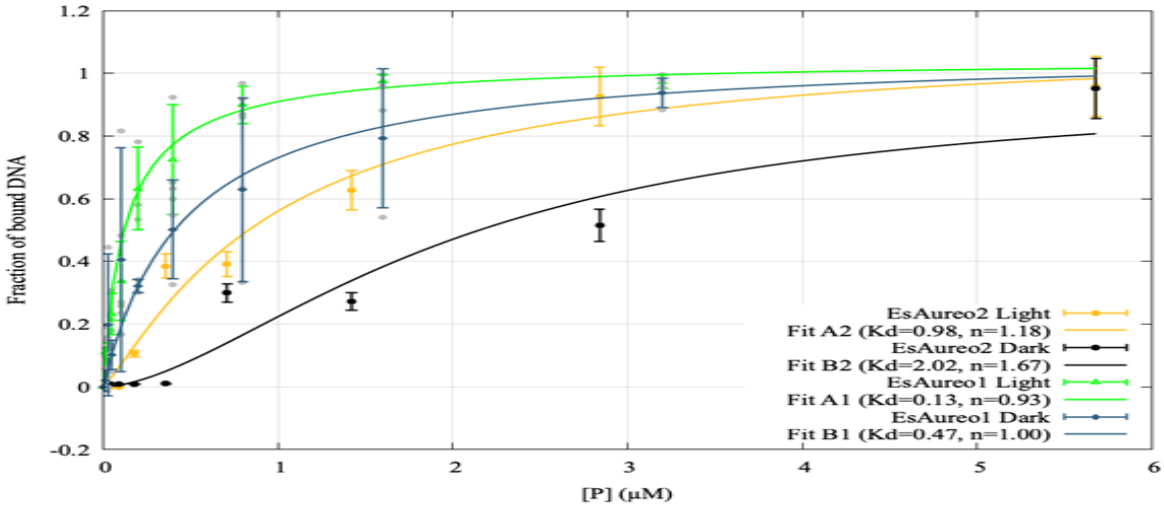


ii iii

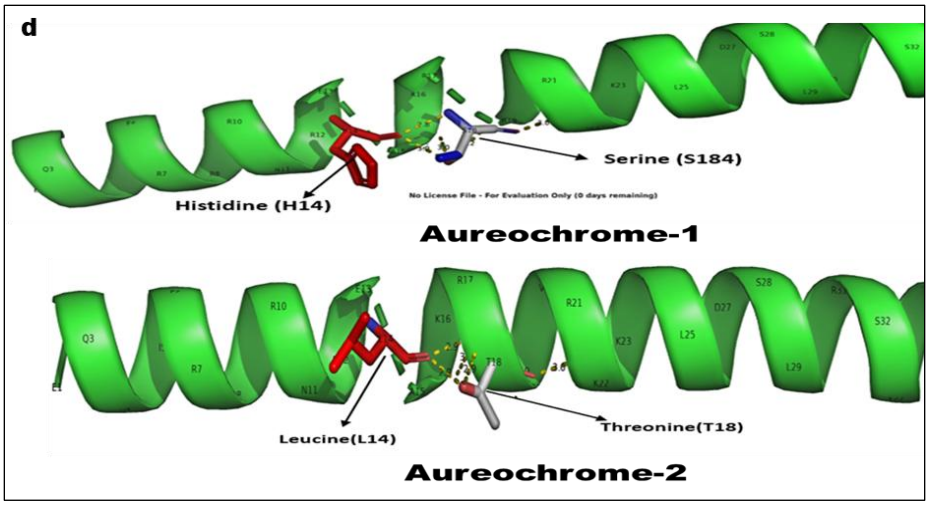
c.



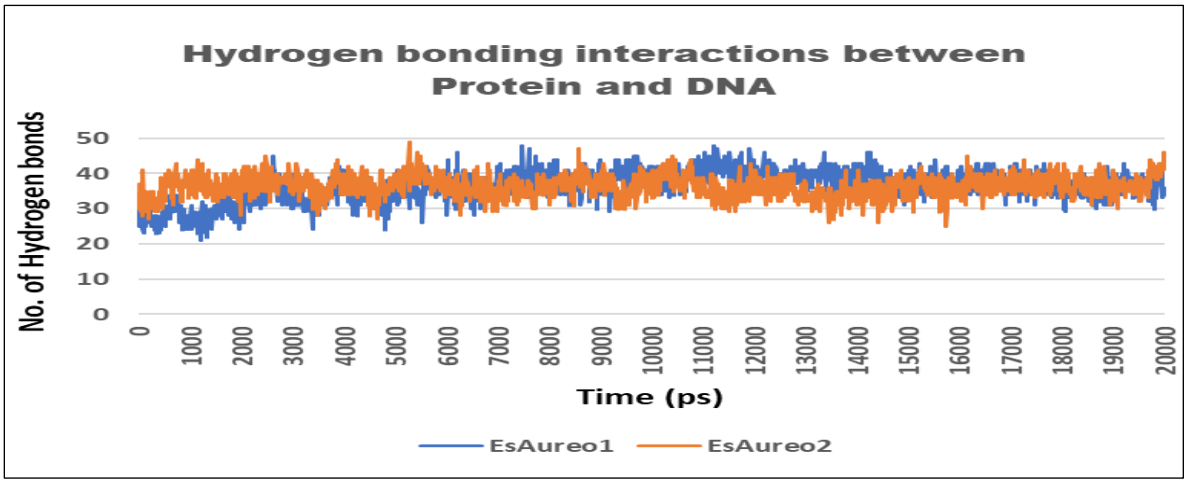
i ii



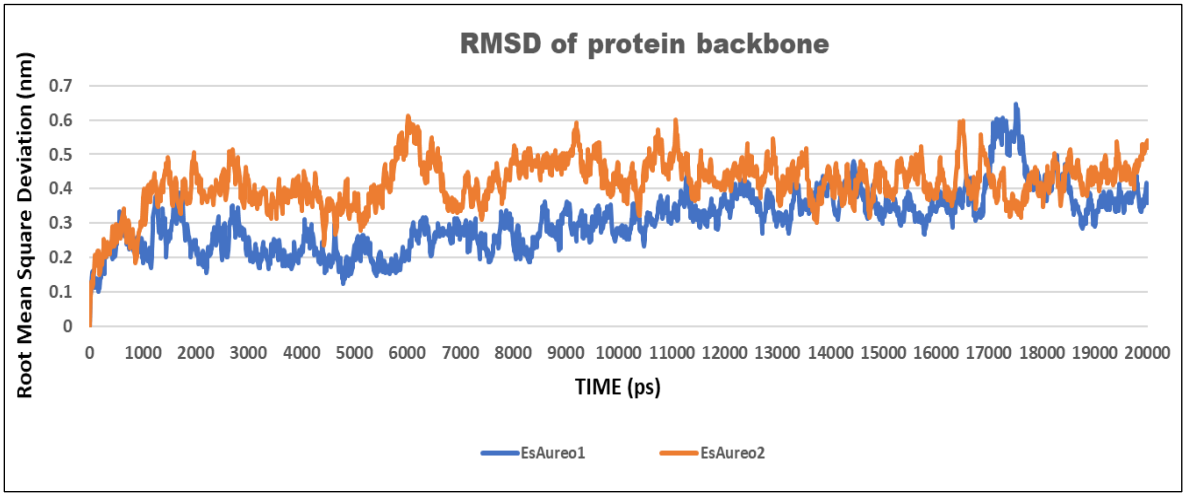
iii


d
Serine (S184)
Histidine (H14)
Aureochrome-1
Threonine(T18)
Leucine(L14)
Aureochrome-2

i


Hydrogen bonding interactions between Protein and DNA
No. of Hydrogen bonds
Time (ps)
EsAureo1
EsAureo2

ii


RMSD of protein backbone
Root Mean Square Deviation (nm)
TIME (ps)
EsAureo1
EsAureo2

iii

Figure-9: a.i. Interface characteristics with hydrogen bonding details of *Es*Aureo2 protein with DNA and FMN. ii. Pair wise alignment of *Es*Aureo1 and *Es*Aureo2. iii. Protein sequence of the *Es*Aureo2 obtained by site directed mutagenesis. b.i. Modelled and docked *Es*Aureo2 protein. ii. UV-Vis spectra and iii. SDS gel image of *Es*Aureo2 protein.

Modelling of *Es*Aureo5 revealed extensive disordered and flexible region flanking the LOV sensor. Hence, only the LOV domain of *Es*Aureo5 protein was cloned [**Figure-10.a**] and validated through sequencing [**Figure-10.b**]. The UV-Vis spectroscopic analysis of the homogeneously purified protein [**Figure-10.d**] exhibited mild triplet peak of absorbance around 447nm [**Figure-10.c**] that does not disappeared upon light exposure.

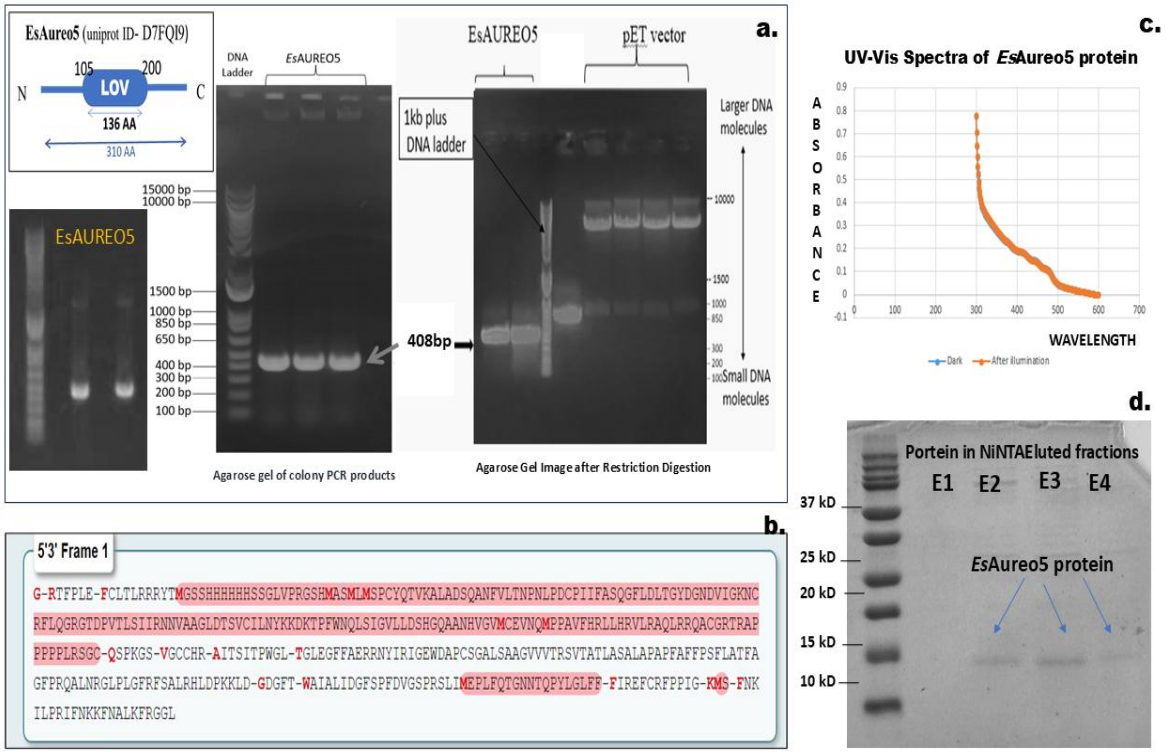


Figure-10: a. Cloning b. sequence validation c. UV-Vis spectra d. SDS gel image of *Es*Aureo5

MSA [**Figure-6.a**] indicated substitution of four of the eleven conserved residues critical for FMN binding. This justified the UV-Vis spectrum of *Es*Aureo5 protein. Presence of bulky methionine in the FMN binding region created steric hindrance leading to improper FMN binding and restriction in the mobility of isoalloxazine ring of FMN. As *Es*Aureo5 protein failed to exhibit proper photochemistry, fluorescence intensity and quantum yield

of this protein was also assessed to verify if the energy gets lost as fluorescence [**Figure-11**].

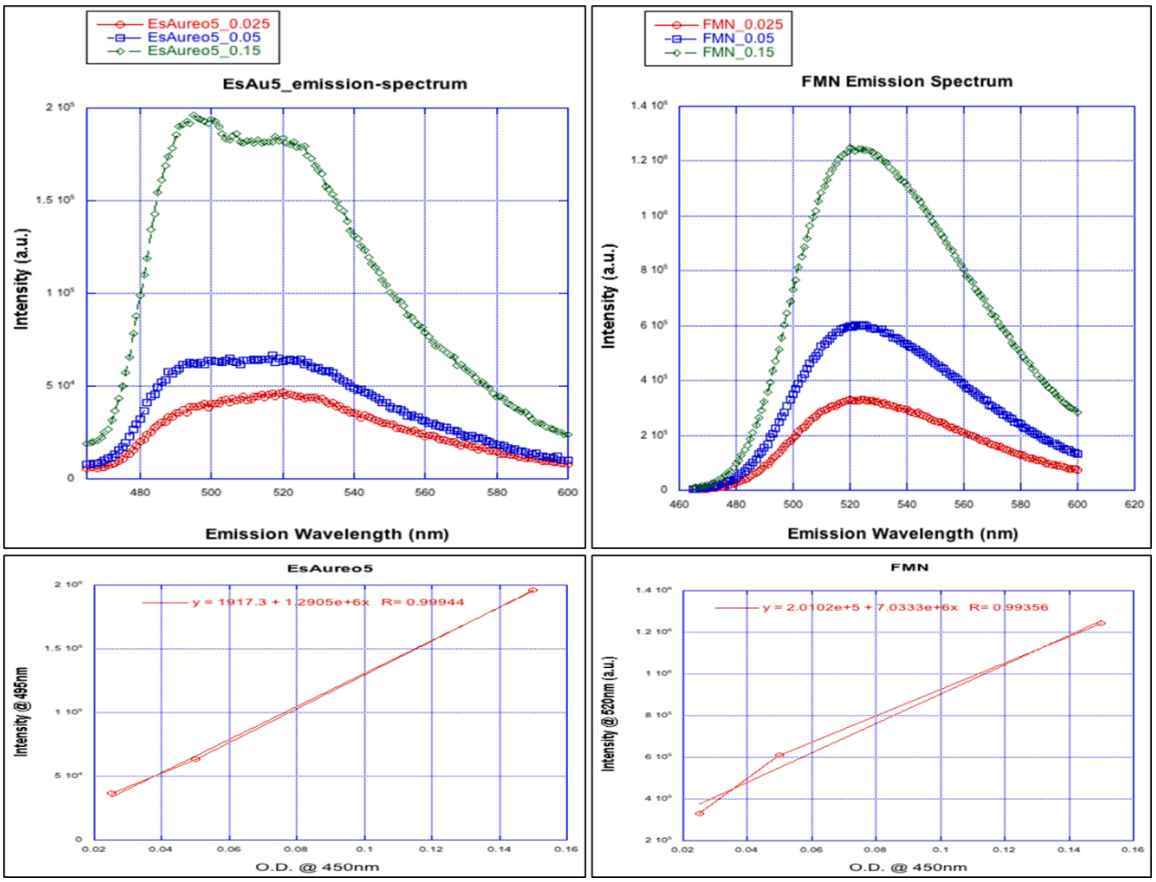


Figure-11: Fluorescence intensity measurement, quantum yield calculation of *Es*Aureo5 protein and free FMN

The measured fluorescence was 0.047, which means 4.77 participates in a photochemical event like-fluorescence for every 100 photons absorbed. This is less than a typical fluorophore's quantum yield that ranges from 0.2 to 0.4.

[A manuscript containing some part of the work done in objective-1 is under revision in a peer reviewed journal and is available online in Biorixiv- "Decoding Sequence-Structure-Function-Evolution of basic Leucine Zippers of Aureochromes from Heterokont Algae" –Khamaru M, Bose D, Deb A. Preprint Biorixiv doi: /10.1101/2022.05.19.492614]

➢ Objective-2:

Understanding combinatorial diversity of aureochrome bZIPs through genome wide screening in *Ectocarpous* sp.

bZIPs bind to DNA as dimers[2728] and forms genome-wide dimer mediated regulatory network[29]. Activation of such gene regulatory network happens not only by interaction with the DNA but through finding of the compatible monomeric partners. Variety of dimers widens the substrate DNA repertoire for the execution of diverse functionalities. It is therefore important to determine the dimerization properties of aureochrome-bZIPs. This study aims to explore the combinatorial diversity of the dimers formed by aureochrome-bZIPs. Screening through the genome of *Ectocarpus siliculosus* revealed four aueochrome-bZIPs and ten non-aureochome bZIPs [**Table-1.a, 1.b**] that offers various combinations for dimerization. To determine if aureochrome-bZIP monomers dimerize among themselves or interact with other bZIPs, 46 dimers were modeled by dimerizing monomeric bZIPs in all possible combinations [**Figure-12**].

Uniprot ID	Name	Sequence
D7FMH8	EsAureo1	EEQRIERRERNREHAKRSRVRKKFLLDSLQRSVDAIQAENESLKGSIVGSLGERGREL
D8LE91	EsAureo2	EEQAKKRRERNRVLARRTRLRKKFFFQSLQQQVARLQRENERLKGIVTTRCPDSVGEIL
D8LT99	EsAureo3	EQQRLDRRERNREHAKRSRVRKKFLLESLQKSVTSLQEENEKLRGAIRSNLGPEEAKEL
D7FRW8	EsAureo4	EQQKLERRERNREHAKRSRIRKKFMLECLQEQLLAMRKQNLALRQVVKEHMPDEASTV

a.

Uniprot ID	Gene name	Sequence
D7FRU4	Esi_0219_0040	KRARREERLMKNREAANRSRVKRKEVLSELENRADTLSKSLAASRDETASLKQEIASLREQNSFLRGMLSAHGSAI
D8LD64	Esi_0113_0081	MDSKRRRRLELNRKAAKESRRRKKMRIEELGRSVVFLTRENQELREQNEIFRQMVASEVSPDGKIGFNTKFRAENA
D7G007	Esi_0039_0117	KEERAKQNRDRNREHARNTRLRKKAFVEELKKQVEELVSGREREEMDQRAEARRVASLREIRLQVLLTFLEYRGRC
D7FPU9	Esi_0199_0057	RKAVAADRKARNRERNRLHARKTRQRKKQHVQELQLKIWGLRQERQRLLQLSTDHRTAGVLLALSCSPKQCQEGPA
D8LF84	Esi_0145_0005	KEVRKQQRMLRNRESAALSRKRKSDRIGELEIQVEALQEENRRLRQRIDRKEDDGRDAGNGNGNGRVFNAGEMTAS
D7G5J2	Esi_0065_0095	MESKRLKRLEKNRESARECRRRKKEHKEKLEAHLASLEEENLNLRLQLRVGDEAEDAENAEILEIKKGLHDQVARG
D7FR79	Esi_0211_0027	GDERRQKRLARNRESARQSRRRKKEHLELLEEKSLGELERHLAAMDASVAASSSGGG
D7FMN5	Esi_0017_0151	MDQRRSRRLERNRASARVRRQRKKNMAELYEKEVARLEACIATLRDHVWGGGE
D8LQU1	Esi_0060_0098	AEIRLERSREQNKQSSKKARIRRKGEEESLKHEIESLQVGQAASAWEIFVTPIHAEPVEGCPALSQDCSCDFYGSL
D7FYX6	Esi_0035_0131	QAWKKQQRMIRNRESAALSRKRKRDKIESLEEQVARLAEENRGLRHRLAKYEASPQQARYKYARQTATTPRSGAPR

b.

Table-1: Sequence of a. aureochrome-bZIPs b. non-aureochrome-bZIPs found in the genome of *Ectocarpus siliculosus*

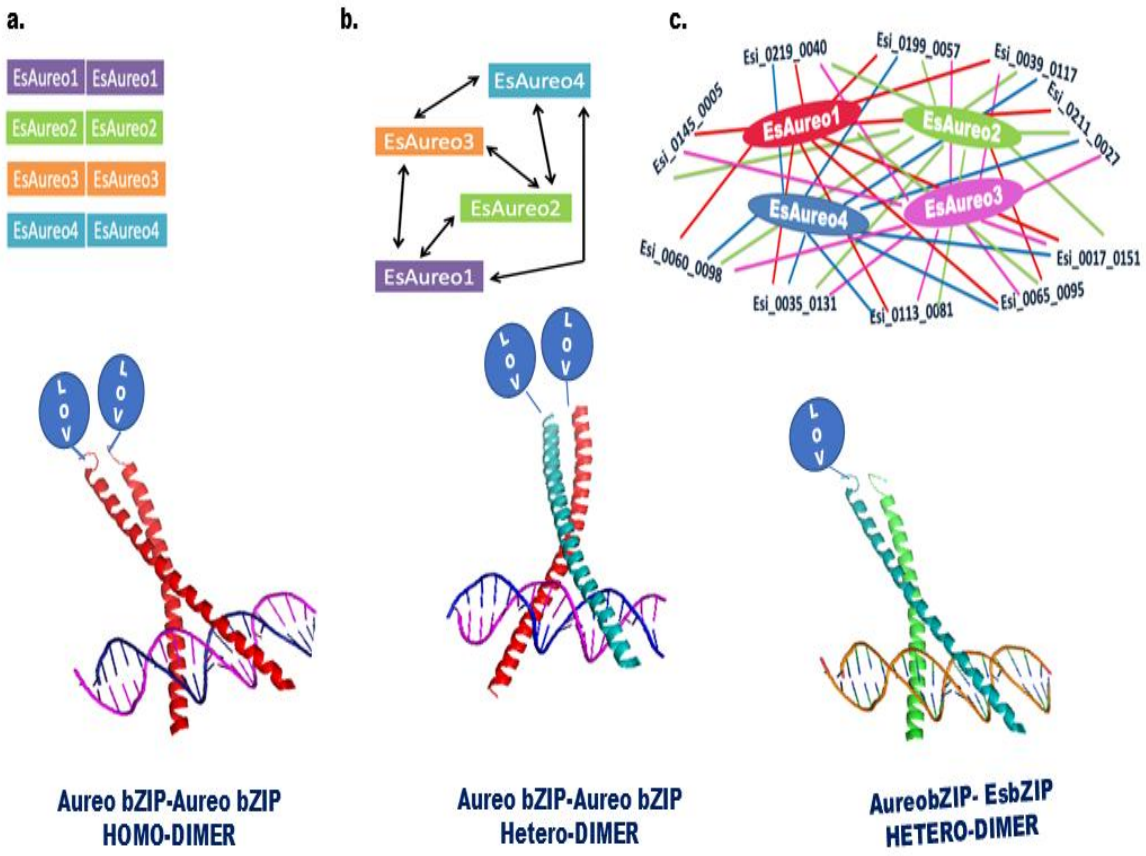


Figure-12: Schematic representation of the homo/hetero-dimerization possibilities. a. *Es*Aureo homodimers b. Heterodimers composed of different subtypes of aureochrome-bZIP monomers c. Heterodimers of *Es*Aaureo-bZIP with other *Es*bZIP monomers

Interface of all the modeled homo/heterodimers were assessed based on the number of saltbridges/ hydrogen-bonds and gain in solvation energy [**Table-2**].

Protein A	Protein B	Template	RMS value	Interface Area (A2)	Delta G	P-value	H-bond	salt bridge
EsAureo1	Esi_0219_0040	2wt7.1	0.083	871	-17.2	0.04	1 A:GLU 39[OE1] 3.45 B:ARG 45[NH1] 2 A:ASN 40[OD1] 3.82 B:SER 40[OG] 3 A:ASN 40[OD1] 3.76 B:SER 44[OG] 4 A:GLU 53[OE1] 3.15 B:ARG 59[NH1]	1 A:ARG 54[NE] 3.58 B:GLU 54[OE1] 2 A:GLU 39[OE1] 3.45 B:ARG 45[NH1] 3 A:GLU 53[OE1] 3.15 B:ARG 59[NH1]
	Esi_0113_0081	5vpe.1	0.089	876.4	-16.3	0.06	1 A:ASN 40[ND2] 2.88 B:ASN 41[OD1] 2 A:GLN 30[OE1] 2.75 B:ARG 26[NH2] 3 A:ILE 36[O] 3.10 B:ASN 41[ND2] 4 A:SER 46[OG] 2.92 B:ASN 48[ND2]	NA
	Esi_0039_0117	5vpe.1	0.095	515	-11	0.04	1 A:GLU 39[OE1] 2.75 B:ARG 41[NH2]	1 A:GLU 39[OE1] 2.75 B:ARG 41[NH2]
	Esi_0199_0057	5vpe.1	0.083	727.1	-12.6	0.08	1 A:ARG 54[NH1] 3.77 B:THR 53[O] 2 A:GLU 39[OE1] 2.79 B:ARG 45[NH2]	1 A:ARG 54[NH1] 3.77 B:THR 53[O] 2 A:GLU 39[OE1] 2.79 B:ARG 45[NH2]
	Esi_0145_0005	5vpe.1	0.084	853.5	-16.1	0.04	1 A:ASN 40[ND2] 2.79 B:ASN 41[OD1] 2 A:GLN 30[OE1] 2.76 B:ARG 26[NH2] 3 A:ILE 36[O] 3.11 B:ASN 41[ND2]	1 A:ARG 54[NE] 3.98 B:ASP 54[OD2] 2 A:ARG 54[NH1] 3.86 B:ASP 54[OD2] 3 A:ARG 54[NH2] 3.18 B:ASP 54[OD2]
	Esi_0065_0095	5vpe.1	0.081	787.5	-14.2	0.07	1 A:ASN 40[ND2] 3.80 B:GLU 40[OE1] 2 A:ASN 40[ND2] 2.86 B:ASN 41[OD1] 3 A:ILE 36[O] 3.15 B:ASN 41[ND2] 4 A:GLU 39[OE1] 3.02 B:ARG 45[NH1]	1 A:GLU 39[OE1] 3.02 B:ARG 45[NH1]
	Esi_0211_0027	5vpe	0.083	640.8	-15.5	0.02	NA	NA
	Esi_0017_0151	5vpe	0.08	623.1	-13.1	0.05	1 A:GLU 39[OE1] 3.04 B:ARG 45[NH1]	1 A:GLU 39[OE1] 3.04 B:ARG 45[NH1]
	Esi_0060_0098	5vpe	0.081	465.6	-9.7	0.07	NA	NA
	Esi_0035_0131	5vpe	0.129	856.9	-17.2	0.05	1 A:ASN 40[ND2] 3.76 B:GLU 40[OE1] 2 A:ASN 40[ND2] 2.77 B:ASN 41[OD1] 3 A:ARG 54[NE] 2.97 B:SER 54[OG] 4 A:ILE 36[O] 3.08 B:ASN 41[ND2] 5 A:GLU 39[OE1] 3.66 B:ASN 41[ND2]	NA
	EsAureo2	5vpe	0.09	747.1	-14.4	0.05	1 A:ASN 40[ND2] 2.84 B:ASN 40[OD1] 2 A:ARG 54[NH1] 3.83 B:ILE 46[O] 3 A:ILE 36[O] 3.13 B:ASN 40[ND2]	1 A:ARG 54[NH1] 3.95 B:VAL 47[OXT]
	EsAureo3	5vpe	0.084	697.7	-15.1	0.03	1 A:ASN 40[ND2] 3.80 B:GLU 39[OE1] 2 A:ASN 40[ND2] 2.79 B:ASN 40[OD1] 3 A:ILE 36[O] 3.12 B:ASN 40[ND2]	NA
	EsAureo4	5vpe	0.359	786	-13.7	0.11	1 B:ASN 40[ND2] 3.12 A:ILE 36[O] 2 B:ARG 44[NH1] 2.93 A:GLU 39[OE1] 3 B:ASN 40[OD1] 2.84 A:ASN 40[ND2] 4 B:VAL 46[O] 3.89 A:ARG 54[NH1] 5 B:LYS 48[O] 3.77 A:ARG 54[NH1]	1 B:ARG 44[NH1] 2.93 A:GLU 39[OE1] 2 B:LYS 48[O] 3.77 A:ARG 54[NH1]

Protein A	Protein B	Template	RMS value	Interface Area (A2)	Delta G	P-value	H-bond	salt bridge
EsAureo2	Esi_0219_0040	5vpe	0.48	731.5	-17.1	0.022	1 B:ASN 40[ND2] 3.15 A:LEU 37[O] 2 B:ASN 40[ND2] 2.96 A:SER 40[OG]	NA
	Esi_0113_0081	5vpe	0.368	723.2	-11.2	0.093	1 A:ASN 40[ND2] 2.86 B:ASN 41[OD1] 2 A:GLN 30[OE1] 2.75 B:ARG 26[NH2] 3 A:LEU 36[O] 3.06 B:ASN 41[ND2] 4 A:LEU 43[O] 3.08 B:ASN 48[ND2]	NA
	Esi_0039_0117	5vpe	0.34	490.1	-11.5	0.029	NA	NA
	Esi_0199_0057	5vpe	0.35	664.6	-15	0.023	1 A:ASN 40[ND2] 3.83 B:LEU 41[O]	NA
	Esi_0145_0005	5vpe	0.037	750.1	-14.8	0.025	1 A:ASN 40[ND2] 2.80 B:ASN 41[OD1] 2 A:GLN 30[OE1] 2.74 B:ARG 26[NH2] 3 A:LEU 36[O] 3.05 B:ASN 41[ND2]	NA
	Esi_0065_0095	5vpe	0.352	736.3	-12.9	0.079	1 B:GLN 32[NE2] 3.14 A:GLU 38[OE2] 2 B:ASN 40[ND2] 3.04 A:LEU 37[O] 3 B:GLU 39[OE2] 3.23 A:ARG 45[NH1] 4 B:ASN 40[OD1] 2.83 A:ASN 41[ND2]	1 B:GLU 39[OE2] 3.23 A:ARG 45[NH1]
	Esi_0211_0027	5vpe	0.359	646.5	-15.1	0.02	1 A:VAL 33[N] 3.66 B:SER 34[OG]	NA
	Esi_0017_0151	1ci6	0.1	604.1	-11.9	0.114	1 B:GLN 37[NE2] 3.67 A:GLU 33[OE1] 2 B:ARG 50[NH1] 3.36 A:ASP 46[O]	1 B:ARG 50[NH1] 3.36 A:ASP 46[O]
	Esi_0060_0098	1ci6	0.117	453.6	-8.3	0.165	1 B:GLN 37[NE2] 3.66 A:GLU 33[OE1] 2 B:ARG 35[NH1] 3.69 A:GLN 38[OE1]	NA
	Esi_0035_0131	5VPE	0.068	707.8	-14.7	0.041	1 A:ASN 40[ND2] 2.81 B:ASN 41[OD1] 2 A:LEU 36[O] 3.07 B:ASN 41[ND2]	NA
	EsAureo1	5vpe	0.09	747.1	-14.4	0.052	1 A:ASN 40[ND2] 2.84 B:ASN 40[OD1] 2 A:ARG 54[NH1] 3.83 B:ILE 46[O] 3 A:ILE 36[O] 3.13 B:ASN 40[ND2]	1 A:ARG 54[NH1] 3.95 B:VAL 47[OXT]
	EsAureo3	5vpe	0.356	742	-14.7	0.033	1 B:ASN 40[ND2] 3.04 A:LEU 36[O] 2 B:GLU 39[OE2] 3.20 A:ARG 44[NH1] 3 B:ASN 40[OD1] 2.81 A:ASN 40[ND2]	1 B:GLU 39[OE2] 3.97 A:ARG 44[NE] 2 B:GLU 39[OE2] 3.20 A:ARG 44[NH1]
	EsAureo4	1ci6	0.106	798.1	-15	0.121	1 A:ARG 37[NH1] 2.86 B:GLN 32[OE1] 2 A:ASN 40[ND2] 3.34 B:LEU 36[O] 3 A:ASN 40[ND2] 3.42 B:GLU 39[OE2] 4 A:ASN 40[OD1] 2.84 B:ASN 40[ND2]	NA

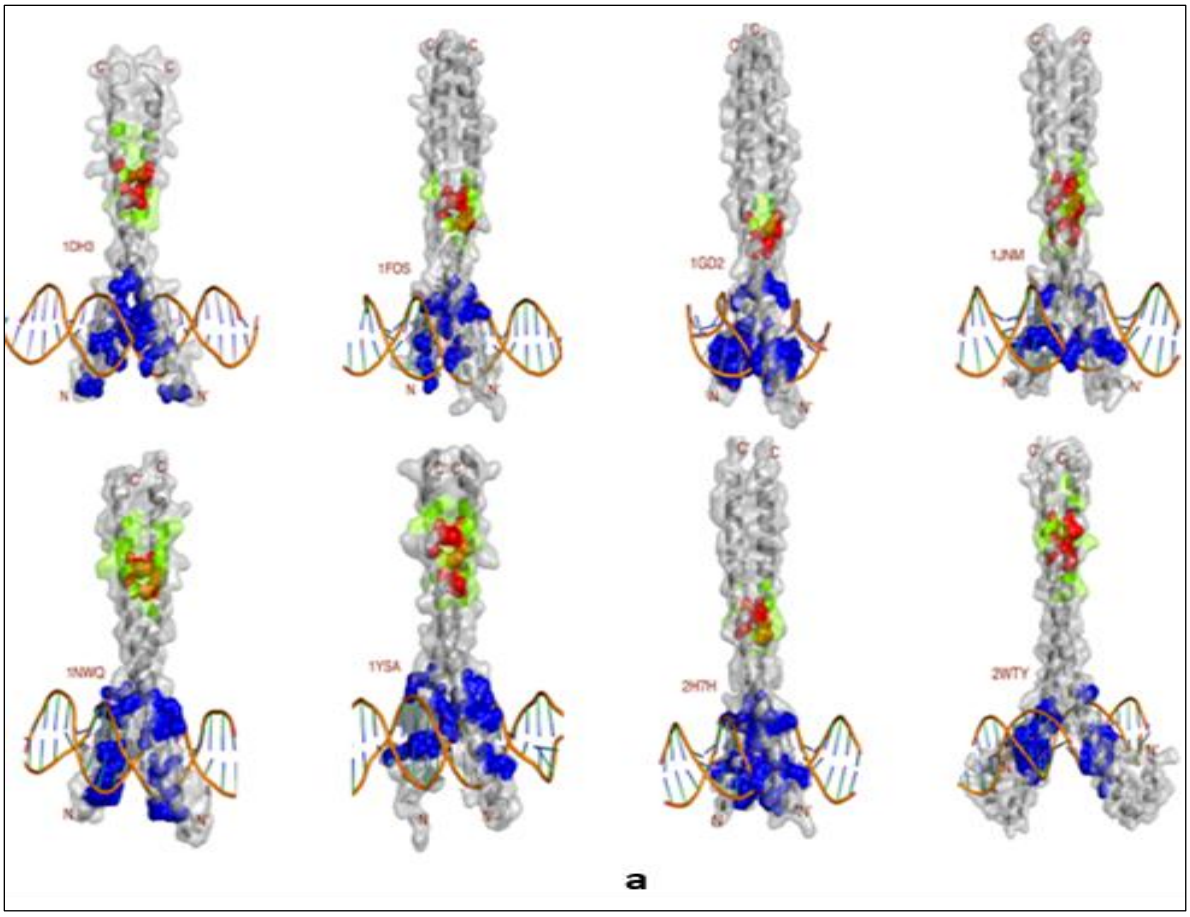
Protein A	Protein B	Template	RMS value	Interface Area (A2)	Delta G	P-value	H-bond	salt bridge
EsAureo3	Esi_0219_0040	1a02	0.113	747.6	-14.9	0.048	1 D:ASN 40[ND2] 3.66 E:SER 40[OG] 2 D:ASN 50[ND2] 3.60 E:THR 48[O] 3 D:GLN 37[OE1] 3.18 E:ARG 33[NH1] 4 D:LEU 43[O] 3.85 E:THR 48[OG1]	NA
	Esi_0113_0081	5vpe	0.08	817.3	-12.4	0.089	1 A:ASN 40[ND2] 2.72 B:ASN 41[OD1] 2 A:ARG 44[NH1] 2.53 B:GLU 40[OE1] 3 A:ASN 50[ND2] 2.90 B:ASN 48[OD1] 4 A:GLN 30[OE1] 2.72 B:ARG 26[NH2] 5 A:LEU 36[O] 3.01 B:ASN 41[ND2]	1 A:ARG 44[NH1] 2.53 B:GLU 40[OE1] 2 A:ARG 44[NH2] 3.97 B:GLU 40[OE1]
	Esi_0039_0117	5vpe	0.076	521.8	-11	0.037	1 A:GLU 39[OE1] 2.78 B:ARG 41[NH2]	1 A:GLU 39[OE1] 2.78 B:ARG 41[NH2]
	Esi_0199_0057	5vpe	0.079	705	-14.6	0.027	1 A:ARG 44[NH1] 3.14 B:GLU 44[OE2] 2 A:ASN 50[ND2] 3.78 B:SER 52[O]	1 A:ARG 44[NE] 3.91 B:GLU 44[OE2] 2 A:ARG 44[NH1] 3.14 B:GLU 44[OE2]
	Esi_0145_0005	5vpe	0.081	803.1	-14	0.035	1 A:ASN 40[ND2] 2.79 B:ASN 41[OD1] 2 A:ARG 44[NH1] 3.20 B:GLU 40[OE2] 3 A:GLN 30[OE1] 2.72 B:ARG 26[NH2] 4 A:LEU 36[O] 3.01 B:ASN 41[ND2]	1 A:ARG 44[NE] 3.95 B:GLU 40[OE2] 2 A:ARG 44[NH1] 3.20 B:GLU 40[OE2]
	Esi_0065_0095	5vpe	0.085	801.2	-13	0.066	1 A:ASN 40[ND2] 3.62 B:GLU 40[OE1] 2 A:ASN 40[ND2] 2.82 B:ASN 41[OD1] 3 A:ARG 44[NH1] 3.51 B:GLU 40[OE1] 4 A:LEU 36[O] 3.05 B:ASN 41[ND2] 5 A:GLU 39[OE1] 2.82 B:ARG 45[NH1]	1 A:ARG 44[NH1] 3.51 B:GLU 40[OE1] 2 A:GLU 39[OE1] 2.82 B:ARG 45[NH1]
	Esi_0211_0027	5vpe	0.077	637.1	-14.9	0.02	1 A:VAL 33[N] 3.65 B:SER 34[OG] 2 A:ASN 40[OD1] 3.18 B:HIS 41[ND1]	NA
	Esi_0017_0151	5vpe	0.08	643.8	-13.2	0.04	1 A:GLU 39[OE1] 2.97 B:ARG 45[NH1]	1 A:GLU 39[OE1] 2.97 B:ARG 45[NH1]
	Esi_0060_0098	5VPE	0.077	474.3	-9.8	0.059	NA	NA
	Esi_0035_0131	5vpe	0.097	809.5	-14.9	0.043	1 A:ASN 40[ND2] 3.66 B:GLU 40[OE1] 2 A:ASN 40[ND2] 2.72 B:ASN 41[OD1] 3 A:ASN 50[ND2] 3.05 B:TYR 51[O] 4 A:LEU 36[O] 3.06 B:ASN 41[ND2] 5 A:GLU 39[OE1] 3.86 B:ASN 41[ND2]	1 A:ARG 44[NH1] 3.91 B:GLU 40[OE1]
	EsAureo1	5vpe	0.084	697.7	-15.1	0.032	1 A:ASN 40[ND2] 3.80 B:GLU 39[OE1] 2 A:ASN 40[ND2] 2.79 B:ASN 40[OD1] 3 A:ILE 36[O] 3.12 B:ASN 40[ND2]	NA
	EsAureo2	5vpe	0.356	742	-14.7	0.033	1 B:ASN 40[ND2] 3.04 A:LEU 36[O] 2 B:GLU 39[OE2] 3.20 A:ARG 44[NH1] 3 B:ASN 40[OD1] 2.81 A:ASN 40[ND2]	1 B:GLU 39[OE2] 3.97 A:ARG 44[NE] 2 B:GLU 39[OE2] 3.20 A:ARG 44[NH1]
	EsAureo4	5vpe	0.078	765.3	-16	0.041	1 A:GLN 37[N] 3.61 B:MET 36[SD] 2 A:ASN 40[ND2] 2.76 B:ASN 40[OD1] 3 A:LEU 36[O] 3.04 B:ASN 40[ND2] 4 A:GLU 39[OE1] 2.89 B:ARG 44[NH1]	1 A:GLU 39[OE1] 2.89 B:ARG 44[NH1]

Protein A	Protein B	Template	RMS value	Interface Area (A2)	Delta G	P-value	H-bond	salt bridge
EsAureo4	EsAureo1	5vpe	0.359	786	-13.7	0.107	1 B:ASN 40[ND2] 3.12 A:ILE 36[O] 2 B:ARG 44[NH1] 2.93 A:GLU 39[OE1] 3 B:ASN 40[OD1] 2.84 A:ASN 40[ND2] 4 B:VAL 46[O] 3.89 A:ARG 54[NH1] 5 B:LYS 48[O] 3.77 A:ARG 54[NH1]	1 B:ARG 44[NH1] 2.93 A:GLU 39[OE1] 2 B:LYS 48[O] 3.77 A:ARG 54[NH1]
	EsAureo2	1ci6	0.106	798.1	-15	0.121	1 A:ARG 37[NH1] 2.86 B:GLN 32[OE1] 2 A:ASN 40[ND2] 3.34 B:LEU 36[O] 3 A:ASN 40[ND2] 3.42 B:GLU 39[OE2] 4 A:ASN 40[OD1] 2.84 B:ASN 40[ND2]	NA
	EsAureo3	5vpe	0.078	765.3	-16	0.041	1 A:GLN 37[N] 3.61 B:MET 36[SD] 2 A:ASN 40[ND2] 2.76 B:ASN 40[OD1] 3 A:LEU 36[O] 3.04 B:ASN 40[ND2] 4 A:GLU 39[OE1] 2.89 B:ARG 44[NH1]	1 A:GLU 39[OE1] 2.89 B:ARG 44[NH1]
	Esi_0219_0040	1a02	0.105	724.3	-15.3	0.057	1 E:ASN 40[ND2] 3.32 D:LEU 37[O]	NA
	Esi_0113_0081	5vpe	0.083	850.9	-13	0.147	1 A:ARG 44[NH1] 2.61 B:GLU 40[OE2] 2 A:ASN 40[ND2] 2.81 B:ASN 41[OD1] 3 A:GLN 30[OE1] 2.71 B:ARG 26[NH2] 4 A:MET 36[O] 3.04 B:ASN 41[ND2]	1 A:ARG 44[NH1] 2.61 B:GLU 40[OE2]
	Esi_0039_0117	1ci6	0.123	547.8	-9.8	0.157	1 A:ARG 37[NH1] 2.82 B:GLN 33[OE1] 2 A:ASN 40[ND2] 3.32 B:LEU 37[O]	NA
	Esi_0199_0057	1ci6	0.114	729.2	-15.5	0.068	1 A:ASN 40[ND2] 3.35 B:LEU 41[O] 2 A:ASN 40[ND2] 3.44 B:GLU 44[OE2] 3 A:HIS 50[NE2] 3.58 B:ASP 54[O] 4 A:GLN 30[OE1] 3.05 B:HIS 30[NE2] 5 A:GLN 39[OE1] 2.68 B:ARG 45[NE]	1 A:HIS 50[NE2] 3.58 B:ASP 54[O]
	Esi_0145_0005	1a02	0.108	854.6	-15.5	0.057	1 D:ARG 44[NH1] 3.06 E:GLU 40[OE2] 2 D:ASN 40[ND2] 2.83 E:ASN 41[OD1] 3 D:GLN 30[OE1] 2.75 E:ARG 26[NH2] 4 D:MET 36[O] 3.26 E:ASN 41[ND2] 5 D:GLN 39[OE1] 2.85 E:ASN 41[ND2]	1 D:ARG 44[NH1] 3.06 E:GLU 40[OE2]
	Esi_0065_0095	5vpe	0.075	779.1	-15.5	0.063	1 B:ASN 40[ND2] 3.06 A:LEU 37[O] 2 B:ARG 44[NH1] 3.03 A:GLU 40[OE1] 3 B:ASN 40[OD1] 2.79 A:ASN 41[ND2]	1 B:ARG 44[NH1] 3.03 A:GLU 40[OE1]
	Esi_0211_0027	1a02	0.116	742.3	-18.3	0.014	1 E:ASN 40[ND2] 3.39 D:LEU 38[O] 2 E:ARG 44[N] 3.86 D:MET 45[SD]	NA
	Esi_0017_0151	1t2k	0.124	511.2	-12.3	0.059	1 E:MET 36[SD] 3.41 F:GLU 38[N] 2 E:GLN 39[OE1] 3.27 F:ARG 45[NH1]	NA
	Esi_0060_0098	1a02	4.309	611.2	-10	0.112	1 E:ARG 37[NH2] 2.92 D:GLU 33[OE2] 2 E:ASN 40[ND2] 3.28 D:LEU 37[O] 3 E:GLN 39[OE1] 3.02 D:GLN 41[NE2]	1 E:ARG 37[NH2] 2.92 D:GLU 33[OE2]
	Esi_0035_0131	1a02	0.13	821.2	-15.2	0.087	1 D:ASN 40[ND2] 3.75 E:GLU 40[OE1] 2 D:ARG 44[NH1] 3.74 E:GLU 40[OE1] 3 D:ASN 40[ND2] 2.75 E:ASN 41[OD1] 4 D:MET 36[O] 3.21 E:ASN 41[ND2] 5 D:GLN 39[OE1] 2.86 E:ASN 41[ND2]	1 D:ARG 44[NH1] 3.74 E:GLU 40[OE1]

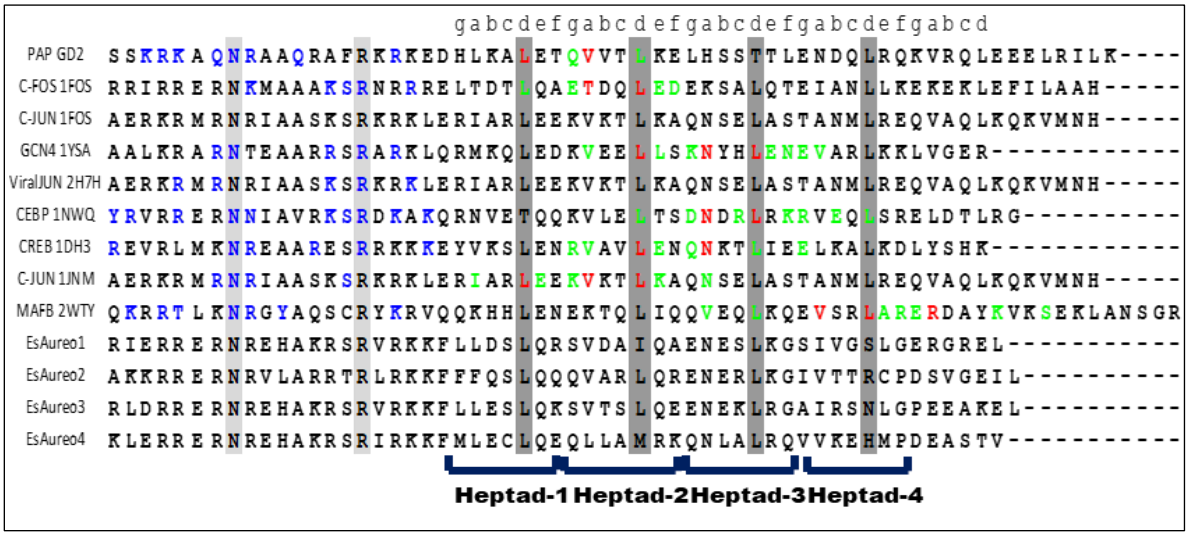
Protein name	Template	RMS	Hydrogen bonding	Salt bridge
EsAureo1 homodimer	6mg3	0.192	1 A:ASN 40[HD21] 2.26 B:ILE 36[O] 2 A:SER 50[OG] 2.71 B:SER 50[OG] 3 A:ARG 54[HE] 2.25 B:SER 50[O]	1 A:ARG 54[NE] 2.69 B:GLU 53[OE1] 2 A:ARG 54[NH1] 3.15 B:GLU 53[OE1] 3 A:ARG 54[NH2] 3.70 B:GLU 53[OE2] 4 A:GLU 53[OE1] 3.15 B:ARG 54[NH2]
EsAureo2 homodimer	6mg2	0.148	1 A:ASN 40[HD21] 2.00 B:LEU 36[O] 2 A:ASN 40[OD1] 1.80 B:ASN 40[HD22]	1 A:LYS 44[NZ] 3.38 B:GLU 39[OE2]
EsAureo3 homodimer	6mg3	0.208	1 A:ASN 40[HD21] 2.11 B:LEU 36[O] 2 A:ASN 40[OD1] 2.02 B:ASN 40[HD21]	NA
EsAureo4 homodimer	2h7h	0.222	1 C:GLN 39[OE1] 2.17 D:ASN 40[HD21]	NA

Table-2: Interface analysis data of modeled homodimers and heterodimers of aureochrome bZIPs

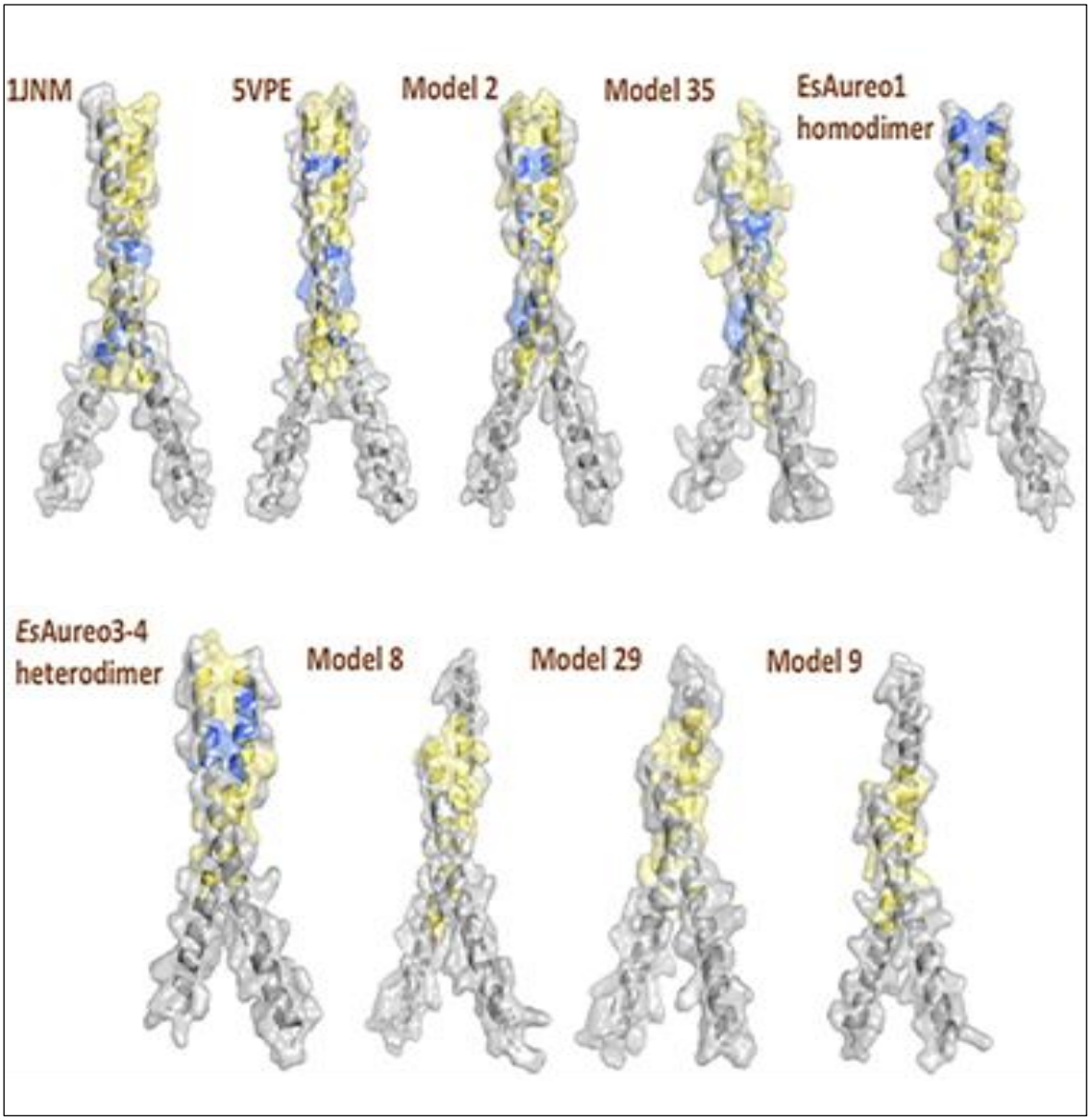
Critical residues for allostery that aid in dimer specificity and stabilization were identified by EC[30] [**Figure-13.a,b**]. The interface residues of the dimers were identified [**Figure-13.c**]. Literature review revealed the residues present at a, g, and e positions of each heptad (a-b-c-d-e-f-g) determine the dimerization specificity[3132]. Asparagine at 'a' increases the chance of homodimerization whereas a charged residue at 'a' inhibits homodimerization. Electrostatic interaction between 'g' of 2nd heptad and 'e' of 3rd heptad located five residues downstream determines dimer stability. Occupancy of identically charged residues at g-e' positions inhibit the formation of homodimers. So, in-depth sequence analysis of the heptad repeats [**Figure-13.d**] of each bZIP monomer was performed to assess the compatibility of the bZIP-monomers.


1DH3
1FOS
1GD2
1JNM
1NWQ
1YSA
2H7H
2WTY

a


g a b c d e f g a b c d e f g a b c d e f g a b c d e f g a b c d
PAP GD2 SSKRKAQNRAAQRAFRKRKEDHLKALETQVVTLKELHSSTTLENDQLRQKVRQLEEELRILK----
C-FOS 1FOS RRIRRERNKMAAAKSRNRRRELTDTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAH-----
C-JUN 1FOS AERKRMRNRIAASKSRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNH-----
GCN4 1YSA AALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER----------
ViralJUN 2H7H AERKRMRNRIAASKSRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNH-----
CEBP 1NWQ YRVRRERNNIAVRKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELDTLRG----------
CREB 1DH3 REVRLMKNREAARESRRKKKEYVKSLENRVAVLENQNKTLIEELKALKDLYSHK----------
C-JUN 1JNM AERKRMRNRIAASKSRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNH-----
MAFB 2WTY QKRRTLKNRGYAQSCRYKRVQQKHHLENEKTQLIQQVEQLKQEVSRLARERDAYKVKSEKLANSGR
EsAureo1 RIERRERNREHAKRSRVRKKFLLDSLQRSVDAIQAENESLKGSIVGSLGERGREL----------
EsAureo2 AKKRRERNRVLARRTRLRKKFFFQSLQQQVARLQRENERLKGIVTTRCPDSVGEIL----------
EsAureo3 RLDRRERNREHAKRSRVRKKFLLESLQKSVTSLQEENEKLRGAIRSNLGPEEAKEL----------
EsAureo4 KLERRERNREHAKRSRIRKKFMLECLQEQLLAMRKQNLALRQVVKEHMPDEASTV----------
Heptad-1 Heptad-2 Heptad-3 Heptad-4

b


1JNM
5VPE
Model 2
Model 35
EsAureo1
homodimer
EsAureo3-4
heterodimer
Model 8
Model 29
Model 9

c.

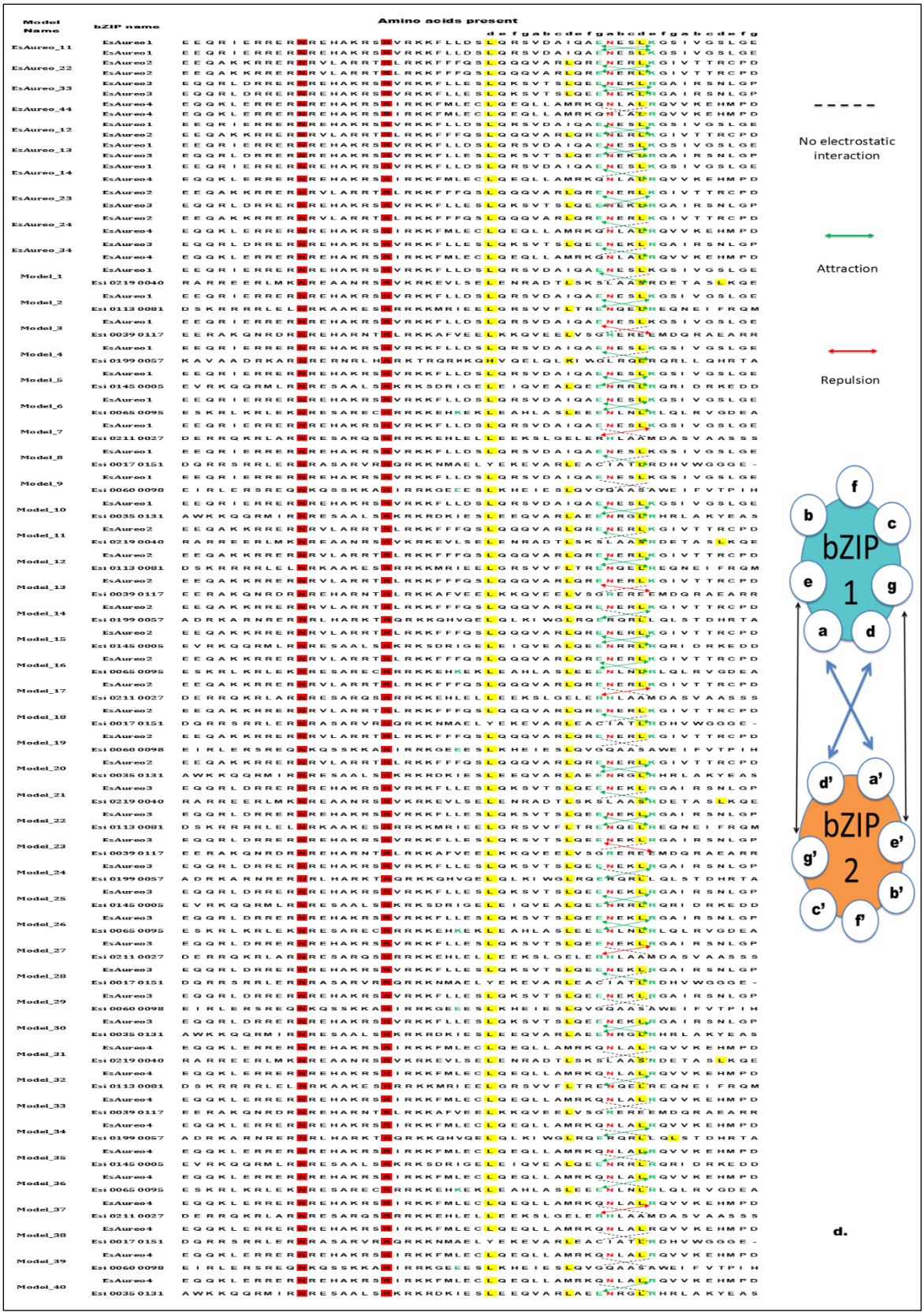

Model Name
bZIP name
Amino acids present
d e f g a b c d e f g a b c d e f g a b c d e f g
No electrostatic interaction
Attraction
Repulsion
bZIP 1
bZIP 2
f
b
c
e
g
a
d
d'
a'
e'
g'
b'
c'
f'
d.

Figure-13: a. Screening of heptad sequences in selected model dimers. The electrostatic attractions and repul-sions between 'g' and 'e' residues are marked with green and red respectively. The non-electrostatic forces including non-polar interactions are marked in black. b. Multiple sequence alignment marking the DNA binding residues of the PDB co-crystal structures in blue and residues with high EC values are colored green/red. c. Residues found important based on eigenvector centrality calculations on eight experimental bZIP-DNA complex structures, selected from PDB are depicted here. Green colour denotes EC value higher than 1.5, Red colour indicates knob like structure (d & a) with the highest EC value, blue colour represents DNA binding residues. The interface residues as obtained from PDBePISA analysis of the selected dimers - 1JNM, 5VPE, Model_2, Model_35, *Es*Aureo1_ homodimer, *Es*Aureo3_4 hetero-dimer, Model_8, Model_29, and Model_9 are highlighted in yellow, whereas residues involved in hydrogen bonding / salt-bridge are highlighted in blue.

The complex network and information theoretic study assessed the compatibility among the protein monomers and the stability of the modelled dimers. ΔE_{SP}, P_{av}, B_r, KL, I_r were the parameters used for the prediction of dimer stabilities [**Figure-14.a**]. While the calculation of the average participation coefficient (P_{av}) and specific difference of energy (ΔE_{SP}) determined the stability of dimers, a higher value of KL divergence implied differences in the structure of dimer and monomer. Ranking of modelled dimers [**Figure-14.a,b**] based on the relative stability revealed highest stability of Model_2 among the heterodimers and *Es*Au1_1 among the homodimers due to higher cumulative value of ΔE_{SP}, P_{av}, B_r, KL, I_r and attractive electrostatic forces implicated between g-e'/g'-e positions.

a.

Model_ID	Ranking	ΔE_{SP}	P_{AV}	KL	I_r	B_r	I_A	ΔG
1jun	0.80125	0.19547	0.17506	2.82295	0.6869	0.73347	923.5	-21
1fos	0.71019	0.18479	0.15265	1.87273	0.67737	0.73712	1188.5	-18.7
5vpe	0.67949	0.14593	0.14243	2.07409	0.68881	0.73155	1171.6	-21
1jnm	0.5988	0.18285	0.1462	1.63791	0.6866	0.73248	1000.6	-14.9
Model_02	0.51011	0.14753	0.13187	1.51889	0.71407	0.7177	876.4	-16.3
Model_37	0.50033	0.15001	0.12907	1.57696	0.72491	0.71588	742.3	-18.3
Model_35	0.49974	0.15496	0.13127	1.53097	0.71117	0.72268	854.6	-15.5
EsAu3_4	0.4924	0.15834	0.12957	1.55752	0.72968	0.7116	765.3	-16
Model_40	0.49171	0.15562	0.13201	1.54025	0.71394	0.72142	821.2	-15.2
EsAu1_1	0.49058	0.14688	0.14307	2.20119	0.72846	0.71108	790.5	-10.3
Model_30	0.49001	0.15848	0.12911	1.57282	0.71989	0.7171	809.5	-14.9
Model_10	0.48321	0.14201	0.11972	1.44944	0.71517	0.71842	856.9	-17.2
Model_26	0.47394	0.15943	0.13143	1.58594	0.72485	0.71357	801.2	-13
Model_05	0.46697	0.14138	0.11852	1.40576	0.71958	0.71453	853.5	-16.1
Model_31	0.46622	0.14887	0.12706	1.81713	0.73084	0.71755	724.3	-15.3
Model_22	0.46278	0.15757	0.13041	1.55107	0.72129	0.71607	817.3	-12.4
Model_36	0.46112	0.1536	0.12162	1.48358	0.73298	0.71184	779.1	-15.5
EsAu4_4	0.4601	0.1383	0.12539	1.91047	0.72944	0.71854	724.3	-15.6
Model_11	0.45816	0.13262	0.1162	1.71205	0.7468	0.70087	731.5	-17.1
Model_32	0.45386	0.14875	0.12659	1.53045	0.72237	0.71696	850.9	-13
Model_07	0.45158	0.14369	0.1258	1.77896	0.71771	0.72028	640.8	-15.5
Model_20	0.45053	0.13804	0.12359	1.87497	0.73667	0.70992	707.8	-14.7
Model_25	0.44763	0.15005	0.12135	1.44667	0.72694	0.71234	803.1	-14
EsAu1_4	0.44508	0.14823	0.12651	1.45623	0.72729	0.71285	786	-13.7
Model_21	0.44232	0.13949	0.12305	1.79067	0.73107	0.71906	747.6	-14.9
EsAu3_3	0.4419	0.12663	0.12371	1.88449	0.74875	0.69462	703.3	-13.7
EsAu2_3	0.43583	0.1428	0.12268	1.47127	0.73703	0.70726	742	-14.7
Model_06	0.42709	0.14085	0.11664	1.43219	0.72503	0.71329	787.5	-14.2
Model_01	0.4204	0.13102	0.10523	1.21647	0.71645	0.72306	871	-17.2
EsAu2_4	0.41252	0.12243	0.11269	1.67723	0.75648	0.70135	798.1	-15
Model_04	0.41228	0.1427	0.12333	1.47358	0.73055	0.71019	727.1	-12.6
Model_34	0.40971	0.12325	0.11014	1.70616	0.75198	0.70162	729.2	-15.5
EsAu1_2	0.40914	0.1334	0.11805	1.39643	0.73604	0.70697	741.1	-14.4
Model_16	0.40704	0.14604	0.11967	1.41375	0.7411	0.70618	736.3	-12.9
Model_24	0.402	0.139	0.11549	1.37754	0.7367	0.70872	705	-14.6
EsAu1_3	0.38369	0.12876	0.1083	1.34515	0.73183	0.70978	697.7	-15.1
Model_15	0.38274	0.12772	0.10974	1.32013	0.74016	0.7078	750.1	-14.8
Model_17	0.38173	0.13118	0.11231	1.4095	0.74464	0.70495	646.5	-15.1
Model_18	0.37469	0.12028	0.11775	1.65727	0.77032	0.67854	604.1	-11.9
EsAu2_2	0.3742	0.1289	0.11181	1.33022	0.75662	0.69072	660	-13.7
Model_27	0.3713	0.1292	0.1091	1.37825	0.74248	0.7047	637.1	-14.9
Model_28	0.35921	0.13291	0.11045	1.35494	0.74664	0.70192	643.8	-13.2
Model_39	0.35354	0.14104	0.11468	1.49001	0.74884	0.69644	611.2	-10
Model_14	0.3492	0.12555	0.10337	1.2867	0.74707	0.70577	664.6	-15
Model_33	0.32703	0.12941	0.1203	1.68782	0.74967	0.70511	547.8	-9.8
Model_12	0.31899	0.12522	0.10521	1.26912	0.73454	0.71305	723.2	-11.2
Model_08	0.31517	0.09889	0.09801	1.45705	0.77032	0.68044	623.1	-13.1
Model_23	0.23966	0.11205	0.09355	1.1323	0.76497	0.69176	521.8	-11
Model_38	0.21594	0.09662	0.08286	1.02095	0.77806	0.68017	511.2	-12.3
Model_03	0.20779	0.10299	0.08704	1.01498	0.77117	0.68804	515	-11
Model_13	0.19542	0.1028	0.0829	1.00612	0.77132	0.69082	490.1	-11.5
Model_29	0.19207	0.09008	0.08124	1.23028	0.78754	0.6718	474.3	-9.8
Model_19	0.18581	0.10751	0.08895	1.3721	0.78177	0.68931	453.6	-8.3
Model_09	0.15802	0.08835	0.07539	1.18139	0.78585	0.67941	465.6	-9.5

b.

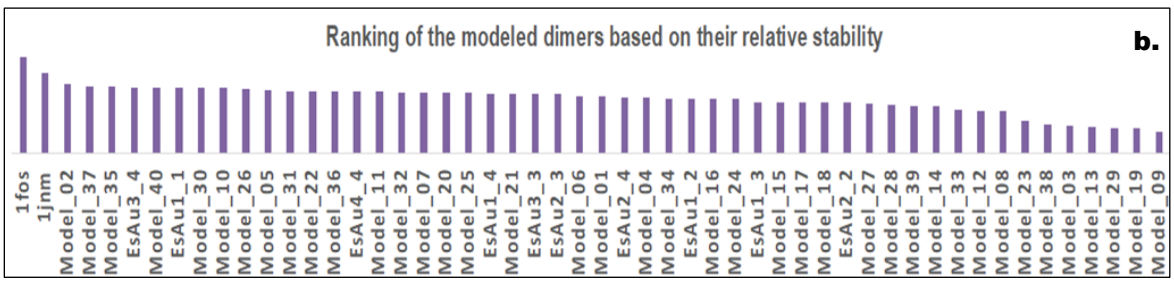


Figure-14: a. Values of specific difference of energy(ΔESP), average participation coefficient (PAV), Kullback-Leibler (KL)divergence, relative information centrality (Ir), relative betweenness centrality (Br), Interface area (IA) and solvation free energy(ΔG) for all possible dimers. b. Ranking of the modelled dimers based on their theoretically calculated stability

Molecular Dynamics Simulations of selected structures validated the mathematical predictions about stability [**Figure-15**].

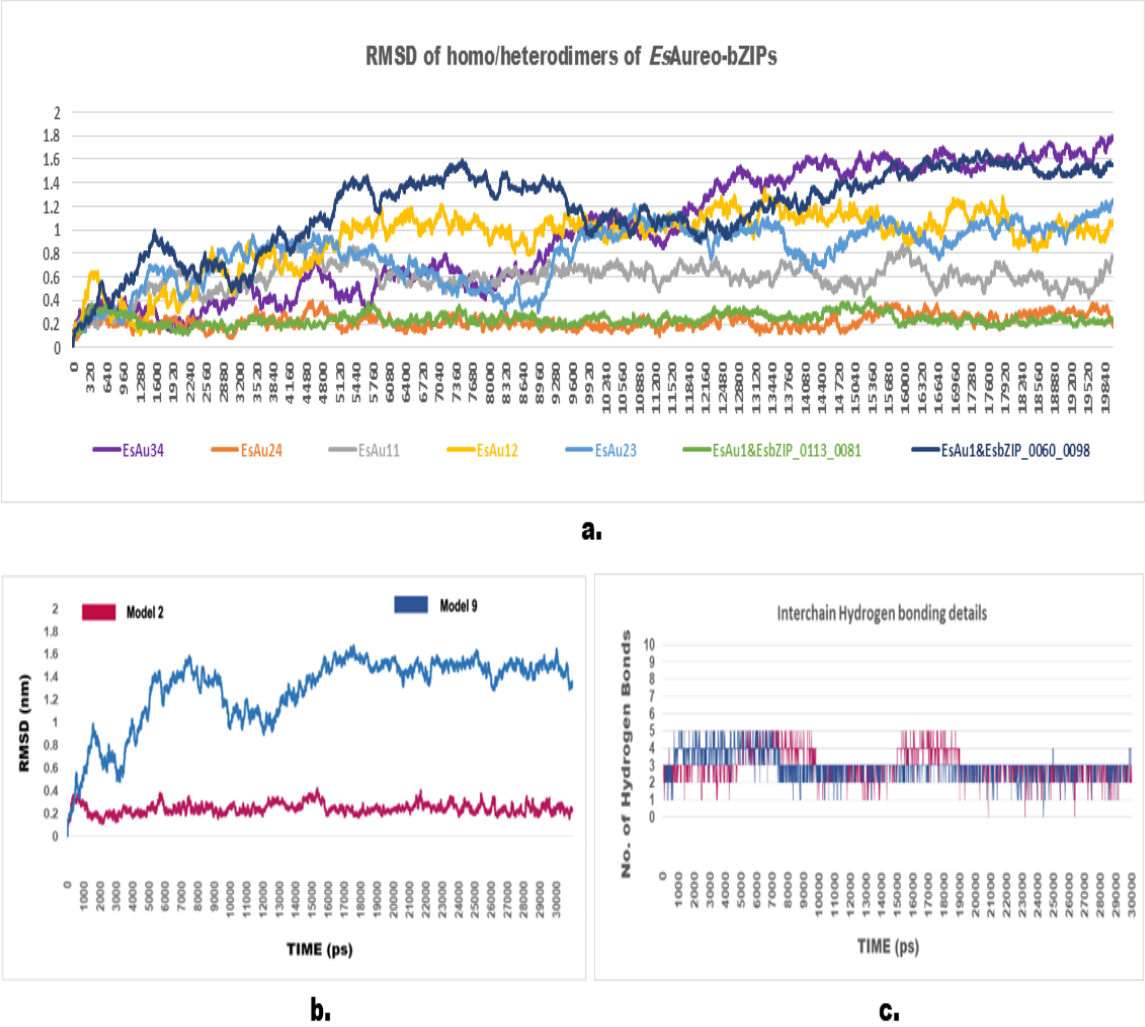


Figure-15: Molecular dynamics (MD) simulations: a. RMSD profile of selected homodimers and heterodimers of Aureochrome-bZIPs. b. RMSD profile of heterodimers (model_2 and model_9) formed by dimerization of an aureochrome-bZIP with a non-aureochrome-bZIP. c. number of inter-chain hydrogen bonds in model_2 and model_9. Distinct variations in RMSD profiles even over a shorter timescale of 30ns predict better stability of model_2.

All findings indicated compatibility of aureochrome-bZIPs with aureo and non-aureo-bZIPs leading to stable dimerization.

- **Objective-3:**

Exploring the potential of aureochromes in plant/algal optogenetics via synthetic studies

Screening and sequence analysis after four rounds of error-prone PCR[26] using different concentrations of $MnCl_2$ and/or $MgCl_2$ [**Table-6.a**] revealed the following randomly incorporated mutations in the LOV and bZIP domain of *Thalassiosira oceanica* (*To*AubZL) [**Table-6.b**].

	$MnCl_2$ and $MgCl_2$ concentrations used during Error-Prone PCR	**No. of Colonies Screened by Colony PCR**	**No. of Constructs containing Mutations in bZIP, Linker, and LOV domains**
Error-Prone PCR Round 1	**0.1 mM $MnCl_2$**	14 (A1–A14)	6 (A1, A3, A6, A9, A12, A13)
	0.1 mM $MnCl_2$ 5 mM $MgCl_2$	Not cloned (PCR product yielded very faint band)	
	0.3 mM $MnCl_2$	2 (B1, B2)	1 (B1)
	0.3 mM $MnCl_2$ 5 mM $MgCl_2$	Not cloned (PCR product yielded very faint band)	
Error-Prone PCR Round 2	**0.1 mM $MnCl_2$**	4 (P1–P4)	0
	0.1 mM $MnCl_2$ 3 mM $MgCl_2$	No colony formed after transformation of ligation products	
Error-Prone PCR Round 3	**0.1 mM $MnCl_2$**	10 (C1–C10)	0
	0.3 mM $MnCl_2$	10 (D1–D10)	1 (D1)
	0.1 mM $MnCl_2$ 2 mM $MgCl_2$	8 (R1–R8)	0
Error-Prone PCR Round 4	**0.1 mM $MnCl_2$**	8 (M1–M8)	0
	0.3 mM $MnCl_2$	5 (N1–N5)	0
	0.1 mM $MnCl_2$ 2 mM $MgCl_2$	14 (S1–S14)	2 (S11)

a.

Construct Name	$MnCl_2$ and $MgCl_2$ Concentration used during Error-Prone PCR	Mutations Incorporated by Error-Prone PCR		Remarks
		bZIP	LOV	
ToAubZL_A1	0.1 mM $MnCl_2$	K318Q	K452E, V499G, A509G	
ToAubZL_A3	0.1 mM $MnCl_2$		A485V, C501G, V503A	No triplet peak
ToAubZL_A6	0.1 mM $MnCl_2$	S314C		
ToAubZL_A9	0.1 mM $MnCl_2$	S314C		
ToAubZL_A12	0.1 mM $MnCl_2$	S324G	D473V, V499R, Q500A, C501V, K502Q, V503S, S504E	No triplet peak
ToAubZL_A13	0.1 mM $MnCl_2$	K298R	V450G, I483M, V499G	
ToAubZL_B1	0.3 mM $MnCl_2$	A311T		
ToAubZL_D1	0.3 mM $MnCl_2$	N307Y, N341S	I414T	No triplet peak
ToAubZL_S11	0.1 mM $MnCl_2$ 2 mM $MgCl_2$	S314C		

b.

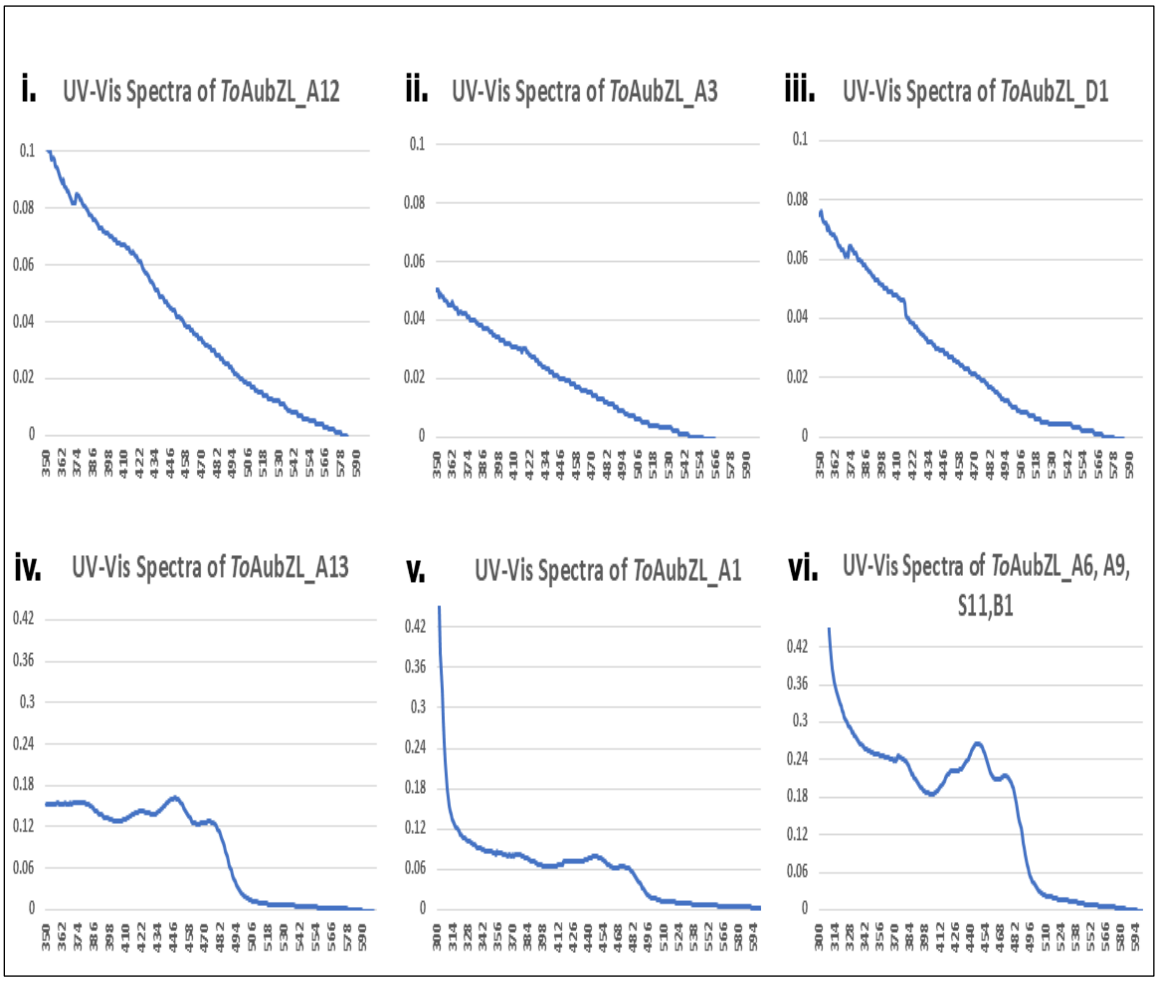
Table-6: a. Details of the 4 rounds of error-prone PCR along with the results of the subsequent cloning into pET28a vector. b. Domain wise documentation of the mutations obtained upon screening and sequencing

Maximum number of mutations were exhibited by ***To*AubZL_A12** concentrated in the LOV domain resulting into loss of the characteristic triplet absorption peak of a LOV sensor around 447nm [**Figure-16.a.i**]. In D473V mutation, the polar side chain of Asp473 helped in keeping the polar contacts/ hydrogen bonding interactions with the side chain and backbone of adjacent Thr475 required for the structural stability of the loop connecting Gβ-Iβ [**Figure-17.b.i**]. Its substitution non-polar Val was thus detrimental. Similarly, inV499R, the substitution of Val499 with Arg at Iβ significantly altered its previous hydrophobic interactions with Ala398 near the A'α, essential for LOV dimerization and signaling. Substitution of the uncharged Val499 with a charged bulky residue have serious impact on the overall structural stability of the LOV core. Another mutation Q500A in the FMN binding pockets of the LOV sensor was vital as the conserved glutamine of this

position undergoes side-chain flipping upon light-driven N5 protonation of FMN, that transmits structural changes to helices bringing alteration in hydrogen bonding. Mutation Q513 was capable of affecting the photochemistry of LOV as found in *As*LOV2. Absence of characteristic triplet absorption peak around 447 was also evident in ***To*AubZL_A3** [**Figure-16.a.ii**] as it possesses three mutations- A485V, C501G, V503A in the core LOV. While Ala485 is located at Hβ, Cys501 and Val503 were located at the distal end of Iβ and in the loop leading to Jα respectively [**Figure-17.d**]. The length of the non-polar side chain of Ala485 was suitable for interaction with nonpolar region of the Tyr496's phenyl ring and Ile493 that was required for maintaining the hydrophobic contact with FMN's isoalloxazine ring. Mutation to Val led to steric clash with Tyr496 and thus hampered the smooth allocation of flavin in the hydrophobic cleft. ***To*AubZL_D1** also failed to show the iconic triplet absorption peak of LOV [**Figure-16.a.iii**] due to a single mutation-I414T. Ile414 residing in close proximity to FMN binding pocket, maintained hydrophobic contacts with the neighboring Val432 and Phe439 needed for stable binding of flavin chromophore[33]. Incidentally, Phe439 was a constituent of fully conserved GXNCRF motif - essential for housing of the flavin chromophore. Its mutation to polar threonine led to loss of those hydrophobic interactions [**Figure-17.a.i,iii**]. *To*AubZL_A6, *To*AubZL_A9, *To*AubZL_B1 and *To*AubZL_S11 in contrast exhibited triplet absorption peak at 447nm having no mutations in LOV [**Figure-16.a.vi**]. Triplet peak of absorption around 447nm was also observed in *To*AubZL_A1 and *To*AubZL_A13 [**Figure-16.a.iv-v**] in spite of the mutations in LOV domain. Though the inherent photochemistry was retained, the rate constant got slightly altered [**Figure-16.b**]. In ***To*AubZL_A13**, the mutation V450G occured at a semi-conserved region [**Figure-18**]. Val450 in Fα remains in the vicinity of flavin chromophore but does not interact with flavin moiety thus exhibit no impact upon mutation to shorter Gly. The other mutation I483M exerted minimal effect like minor alteration in rate constant [**Figure-16.b**] of the photocycle kinetics. Met483 instead of steric clash formation, got flipped towards Fα and made additional hydrophobic contact with Ile457 maintaining descent distance with FMN [**Figure-17.e**]. In **ToAubZL_A1**, mutations K452E and A509G occured at the two flanking ends of the LOV core-A'α and Jα helices respectively [**Figure-17.c.i**]. As the long side chain of Lys452 was buried towards the beta-core, substitution of Lys452 with another polar, charged or even neutral

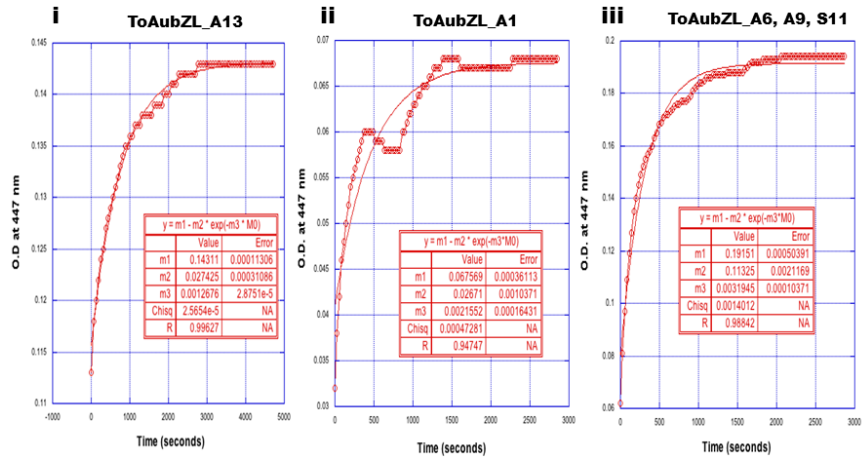
amino acid (like Asn or even Thr as present in some LOV domains) did not affect its wild-type like characteristics. Similarly, the substitution of buried Ala509 of Jα helix with short Gly in A509G was incapable of bringing out alterations in the typical wide-type like behavior.

Val'503 and Val'499 can be designated as mutational hotspots. Val'503 was substituted by alanine and seine in *To*AubZL_A3 and *To*AubZL_A12 respectively. Val'499 was mutated to glycine in both ToAubZL_A1 and ToAubZL_A13 and to positively charged arginine in *To*AubZL_A12.


i. UV-Vis Spectra of ToAubZL_A12
ii. UV-Vis Spectra of ToAubZL_A3
iii. UV-Vis Spectra of ToAubZL_D1
iv. UV-Vis Spectra of ToAubZL_A13
v. UV-Vis Spectra of ToAubZL_A1
vi. UV-Vis Spectra of ToAubZL_A6, A9, S11,B1

a.

Construct Name	Rate Constant (S^{-1})
*To*AubZL_A1	$2.15\times10^{-3}\pm1.64\times10^{-4}$
*To*AubZL_A3	Triplet peak not found
*To*AubZL_A6	$3.19\times10^{-3}\pm1.04\times10^{-4}$
*To*AubZL_A9	$3.19\times10^{-3}\pm1.04\times10^{-4}$
*To*AubZL_A12	Triplet peak not found
*To*AubZL_A13	$1.26\times10^{-3}\pm2.88\times10^{-5}$
*To*AubZL_B1	$3.19\times10^{-3}\pm1.04\times10^{-4}$
*To*AubZL_D1	Triplet peak not found
*To*AubZL_S11	$3.19\times10^{-3}\pm1.04\times10^{-4}$

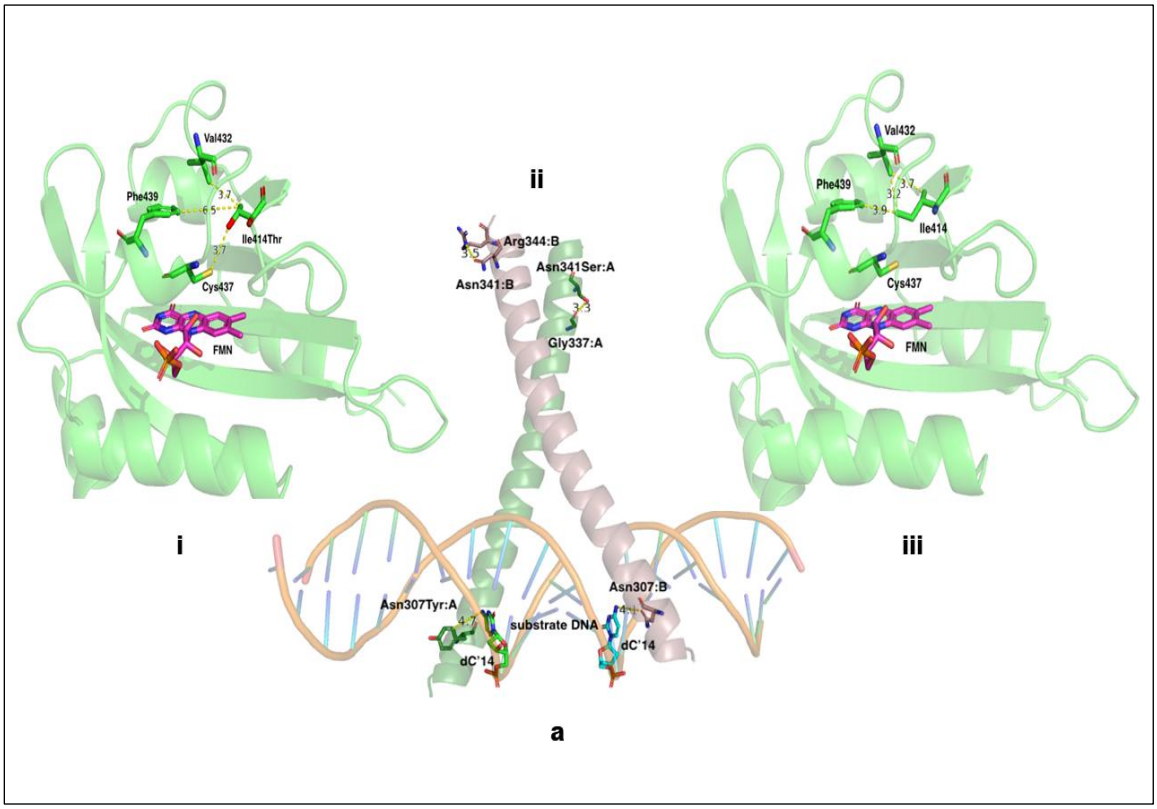


b.

Figure-16: a. UV-Vis spectra b. Rate constant of the photocycle kinetics of the different *To*AubZLvariants

Mutations in the bZIP domain were found in all variants except- *To*AubZL_A3. In ***To*AubZL_A6, *To*AubZL_A9 and *To*AubZL_S11**, S314C mutation occured in a partially conserved position that got mutated to various residues in the different bZIPs. Polar contacts and hydrogen bonds with the substrate DNA were retained [**Figure-17.g**] despite this mutation. In ***To*AubZL_D1** N307Y is the most striking mutation as asparagine, the key residue for bZIP-DNA interaction is otherwise conserved in any bZIP. Asn307 makes base-specific contact with DC'14 (N3) of substrate DNA [**Figure-17.a.ii**]. Though this interaction was lost upon substitution to Tyr307, weak stacking interaction between nitrogenous base and aromatic ring compensated the loss. The homologous Asn235 when mutated to tryptophan in GCN4, the resultant mutant behaved indistinguishably as the wild type bZIP[34] for a specific DNA substrate. Mutation N341S occurred in a poorly conserved region [**Figure-18**] of the dimerization domain. The side chains of Asn341 being exposed to external solvent participated in intra-helical hydrogen-bonding interactions but not a part of dimer interface [**Figure-17.a.ii**]. Asn341's replacement with another polar residue was inconsequential for the retention of intra-helical hydrogen bonding interaction albeit with different residues. ***To*AubZL_B1** possess the mutation-A311T in the signature motif of the DNA binding basic region. Though such a substitution of alanine to threonine was not found in nature, several mutations were naturally found at this particular position in other

bZIPs like- PAP, CEBP [**Figure-18**]. The side chain of A311T by CG made polar contact with the nitrogenous base DT'13 at C5 atom [**Figure-17.f**] adding a bit of specificity to the substrate DNA interaction of Aureo1**.** ***To*AubZL_A13** possess K298R mutation with least impact because the position is beyond the basic region responsible for binding with the substrate DNA. S324G mutation in ***To*AubZL_A12** lies in between the DNA binding basic region and the dimerization domain. Ser324 position is conserved in all Aureo1 homologs but not in other bZIPs [**Figure-18**]. Based on location in a helical turn [**Figure-17.b.ii**], a polar side chain is favorable for its interaction with external solvent. Substitution with a shortest glycine is perfect for the placement in a helical turn as only its polar backbone gets exposed to the solvent due to absence of Cα. In ***To*AubZL_A1**, K318Q mutation represents substitution of a positively charged arginine with an uncharged glutamine that retains polar contact with the phosphate backbone of the substrate DNA, while engaging in hydrogen bonding interaction with the side chain of its neighboring Arg317 [**Figure-17.c.ii**].



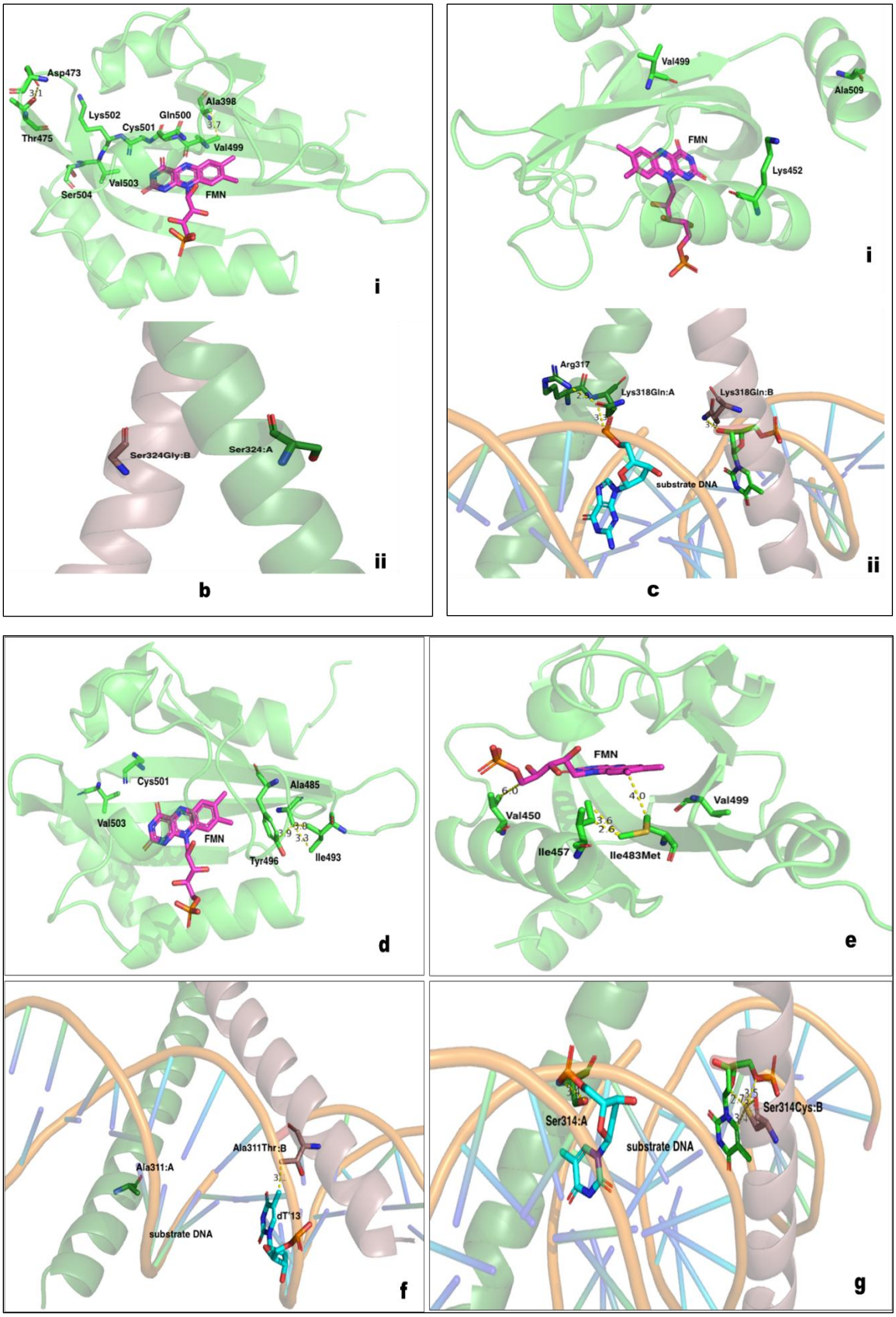

Asp473
Thr475
Lys502
Cys501
Gln500
Ala398
Val499
Val503
Ser504
FMN
i
Ser324Gly:B
Ser324:A
ii
b
Val499
Ala509
FMN
Lys452
i
Arg317
Lys318Gln:A
Lys318Gln:B
substrate DNA
ii
c
Cys501
Ala485
Val503
FMN
Tyr496
Ile493
d
FMN
Val499
Val450
Ile457
Ile483Met
e
Ala311Thr:B
Ala311:A
dT13
substrate DNA
f
Ser314:A
Ser314Cys:B
substrate DNA
g

Figure-17: Important polar and hydrophobic interaction details of a. *To*AubZL_D1, b. *To*AubZL_A12, c. *To*AubZL_A1, d. *To*AubZL_A3, e. *To*AubZL_A13, f. *To*AubZL_B1, g. *To*AubZL_A6,A9,S11.

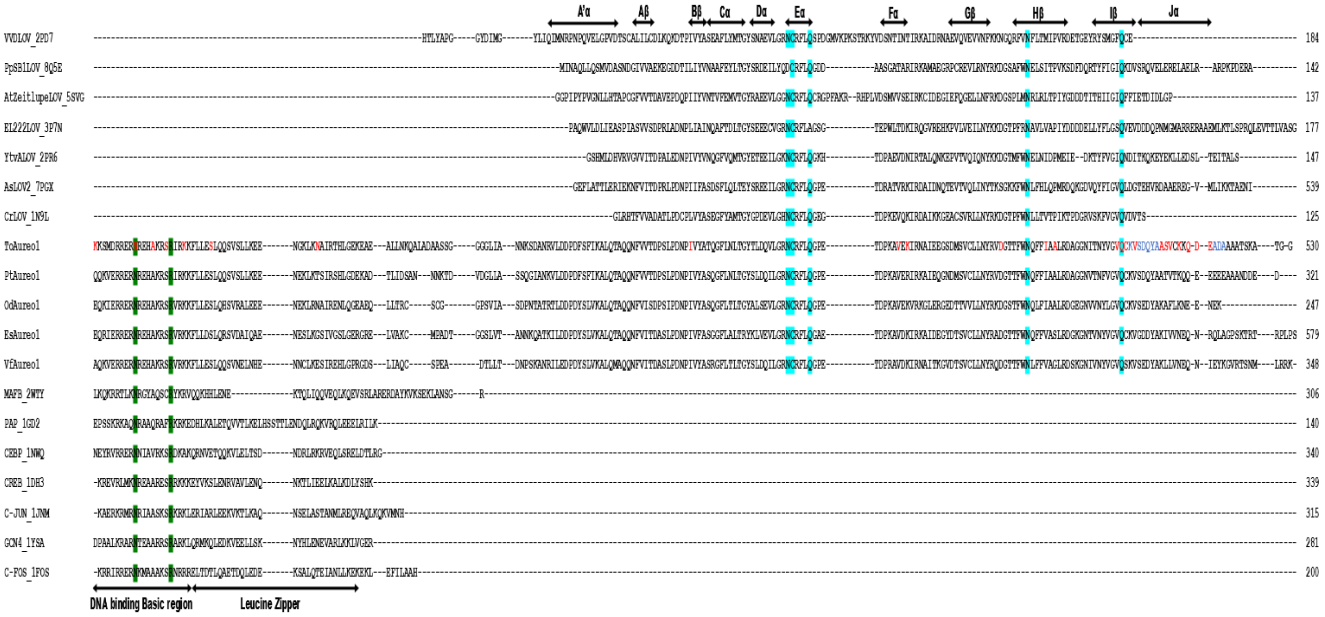


Figure-18: MSA of ToAubZL with other aureochromes and the LOV and bZIP proteins whose crystal structures are available in PDB to determine the residual differences. Residues of ToAubZL that are found to be mutated in the variants of ToAubZL obtained through directed evolution are colored in red/blue. FMN binding residues of LOV domain are colored in cyan and the residues present at the boundary of DNA binding signature motif of bZIP are coloured in green.

Despite the mutations, all variants except *To*AubZL_A3, *To*AubZL_A12 and *To*AubZL_D1 exhibited comparable DNA binding efficiency with two-fold or more difference in binding affinity between light and dark [**Figure-19**] like-in wild type *To*AubZL[10]. Loss of LOV signaling in *To*AubZL_A3, *To*AubZL_A12 and *To*AubZL_D1 inhibitsed their blue light induced DNA binding ability too. They exhibited comparable DNA binding efficiency in dark and light without any significant difference in K_D [**Figure-19**].

Construct Name	Conc. at which 1st band of free DNA appeared in Dark (µM)	Conc. at which 1st band of free DNA appeared in Light (µM)	K_D value		Difference in DNA binding efficiency in Light vs Dark
			Dark	Light	
*To*AubZL_A1	1.44	0.72	0.51µM	0.17µM	Higher in Light
*To*AubZL_A3	0.8	0.8	0.55µM	0.77µM	Negligible difference
*To*AubZL_A6, A9, S11	1.28	0.64	1.29µM	0.53µM	Higher in Light
*To*AubZL_A12	0.52	0.52	0.87µM	0.54µM	Negligible Difference
*To*AubZL_A13	1.2	0.6	0.95µM	0.26µM	Higher in Light
*To*AubZL_B1	1.36	0.68	0.36µM	0.15µM	Higher in Light
*To*AubZL_D1	0.6	0.6	0.47µM	0.47µM	No Difference

Figure-19: DNA binding efficiency of the different variants of *To*AubZL in Light and Dark

CONCLUSION

- **Objective-1**-

1. The basic leucine zipper motif and light oxygen voltage sensor of aureochrome-1 of *Ectocarpus siliculosus* is functionally active and similar to other aureochrome homologs except aureochrome-2.
2. Presence of histidine in the DNA binding N-terminal basic region of aureochrome-bZIP bespoke its light regulated binding to Aureobox-DNA substrate with higher affinity.

- **Objective-2**-

1. Aureochrome-bZIP monomers form stable homodimers as well as inter subtype heterodimers.
2. Aureochrome-bZIPs are compatible with other non-aureochrome-bZIPs and form stable heterodimers by partnering with the non-aureochrome-bZIPs.

- **<u>Objective-3</u>**-

1. The synthetic variants of aureochrome generated by directed evolution maintains the two-fold difference in the affinity of DNA binding between light and dark.
2. They exhibit different photocycle reaction kinetics and DNA binding efficiency that is useful to bring the desired regulation pattern.

SIGNIFICANCE OF THE STUDY

Different aureochrome homologs from various algae have been found to perform different functionalities like- branching in *Vaucheria sp.*[1], cell cycle progression in *Phaeodactylum sp.*[35], lipid accumulation in *Nannochloropsis sp.*[36], optimization of photosynthesis in *Phaeodactylum sp.*[37] under varied light conditions, high light acclimation, photo protection in diatoms[38] etc. Such species-specific functionalities might have arisen from the subtle differences in the amino acid sequences among the various paralogs. Thus, characterization of the different sub-types of aureochrome from *Ectocarpus siliculosus* proved to be useful in revealing the sequence, structural and functional differences among the different Aureochromes.

Aureochrome serves as the major BL receptor in the marine algae under the group 'photosynthetic stramenopiles'. Phototropin, the BL photoreceptor concerned with the high light acclimation of higher plants is completely absent in these aureochrome bearing organisms. It is therefore important to decipher the BL mediated gene regulatory network of aureochrome to unravel the BL dependent developmental mechanism of this lower group under fluctuating light intensities in absence of phototropin. Unlike phototropins, aureochrome doesn't undergo an elaborate signaling cascades via serine threonine kinase. Due to the presence of a DNA binding bZIP motif, it promotes BL dependent gene regulation by directly binding to DNA as homo/heterodimers. To understand how aureochrome by its varied oligomerization properties, binds to various substrates of DNA for bringing diverse gene regulation, the diversity of dimers that can be formed by aureochrome-bZIPs has been determined followed by the assessment of the stability of all the dimers.

The photocycle kinetics as well as the DNA binding ability of aureochromes from different organisms varies. The duration of activation of the signaling state of a photo-switch is crucial as both the short- or long-signaling states are desirable in different optogenetic applications

depending on the context and the specific effector function to be targeted. Efficient binding with highest affinity to a particular DNA substrate is equally important for the proper functioning of an optogenetic tool. So, several synthetic variants with different DNA binding affinity and varying signaling state lifetime must be developed for different purposes based on regulation pattern. Directed evolution has proved to be a promising way to optimize proteins with functions that are not observed in nature[3940]. Out of all the natural aureochromes studied so far (in our lab), *Thallasiosira oceanica* aureochrome (*To*AubZL) has proved to be the most suited candidate to be used as an optogenetic scaffold. Therefore, in this study, engineering of *To*AubZL has been carried out using directed evolutionary strategy for the generation of its synthetic variants. The DNA binding affinity and photochemical property in terms of the rate constant of the adduct state of the variants have been determined in order to obtain the appropriate one that can be used as an opto-genetic tool out of the several synthetic and natural homologs of aureochrome.

FUTURE DIRECTION

Natural light-responsive proteins can be instrumental for the development of the ever-growing suite of opto-genetic tools. In fact, engineering of light-regulated transcription factors is the most adapted strategy in synthetic biology for the development of reversible photo switches. The modularity of the domains, broad spectrum of the on/off kinetics and the association with quantum efficient endogenous chromophores makes aureochrome a suitable candidate to be used in optogenetics. There are instances of use of the LOV sensors as optogenetic tools, but the application of the entire aureochrome in optogenetics is yet to be witnessed. Through the exploration of the coupling preferences, photochemistry and DNA binding efficiency of natural aureochrome homologs and their synthetic variants, this study offers a repertoire of aureochrome that can be used for the development of optogenetic tool according to the requirement, gene target and desired regulation pattern. It can be introduced in the target cells either as a transgene to be expressed under native promoters or as a genetically encoded, synthetic promoter driven recombinant protein. Aureochromes can be employed through the design of orthogonal gene regulatory circuits to promote non-invasive control of the existing regulatory networks.

Aureochrome being sensitive to a particular wavelength of daylight avoids inadvertent activation and interference with the endogenous light driven signaling of plants. Different biological

processes of plants like- photo morphogenesis, biotic and abiotic stress response are controlled by bZIPs. Although bZIPs are classified into different subgroups based on their DNA specificities but many of the plant bZIPs are found to bind at a sequence TGACGTG where aureochrome can bind most efficiently as concluded from this study. Thus, incorporation of aureochrome in plant cell can result in binding of aureochromes to the promoters of different genes involved in the developmental pathways and stress responses as homodimers or by formation of heterodimers with their inherent bZIPs to promote light-mediated tissue-specific regulation of genes.

REFERENCES

1. Takahashi F, Yamagata D, Ishikawa M, et al. AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proc Natl Acad Sci*. 2007;104(49):19625-19630. doi:10.1073/pnas.0707692104
2. Rayko E, Maumus F, Maheswari U, Jabbari K, Bowler C. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol*. 2010;188(1):52-66. doi:10.1111/j.1469-8137.2010.03371.x
3. Matiiv AB, Chekunova EM. Aureochromes – Blue Light Receptors. *Biochem*. 2018;83(6):662-673. doi:10.1134/S0006297918060044
4. Derelle R, On L Opez-Garc Ia P, Ne Timpano H, Moreira D. A Phylogenomic Framework to Study the Diversity and Evolution of Stramenopiles (5Heterokonts). doi:10.1093/molbev/msw168
5. Glantz ST, Carpenter EJ, Melkonian M, et al. Functional and topological diversity of LOV domain photoreceptors. *Proc Natl Acad Sci*. 2016;113(11):E1442-E1451. doi:10.1073/pnas.1509428113
6. Crosson S, Rajagopal S, Moffat K. The LOV domain family: Photoresponsive signaling modules coupled to diverse output domains. *Biochemistry*. Published online 2003. doi:10.1021/bi0269781
7. Rodríguez-Martínez JA, Reinke AW, Bhimsaria D, Keating AE, Ansari AZ. Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. Published online 2017. doi:10.7554/eLife.19272.001

8. Pogenberg V, Consani Textor L, Vanhille L, Holton SJ, Sieweke MH, Wilmanns M. Design of a bZip Transcription Factor with Homo/Heterodimer-Induced DNA-Binding Preference. *Structure*. 2014;22(3):466-477. doi:10.1016/j.str.2013.12.017

9. Peters AF, Marie D, Scornet D, Kloareg B, Cock JM. PROPOSAL OF ECTOCARPUS SILICULOSUS (ECTOCARPALES, PHAEOPHYCEAE) AS A MODEL ORGANISM FOR BROWN ALGAL GENETICS AND GENOMICS1,2. *J Phycol*. 2004;40(6):1079-1088. doi:10.1111/J.1529-8817.2004.04058.X

10. Deb A, Sarkar P, Mitra D. Exploring the Modularity of *Thalassiosira oceanica* Aureochrome1 Photoreceptor-Transcription Factor: Characterization and Component Engineering. Published online March 11, 2025. doi:10.1101/2025.03.07.642083

11. Berman HM, Battistuz T, Bhat TN, et al. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr*. 2002;58(Pt 6 No 1):899-907. doi:10.1107/S0907444902003451

12. Apweiler R, Bairoch A, Wu CH, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2018;46(5):2699-2699. doi:10.1093/NAR/GKY092

13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2

14. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7. doi:10.1038/MSB.2011.75

15. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018;35(6):1547-1549. doi:10.1093/MOLBEV/MSY096

16. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22(22):4673-4680. doi:10.1093/NAR/22.22.4673

17. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303. doi:10.1093/NAR/GKY427

18. Meng EC, Goddard TD, Pettersen EF, et al. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci*. 2023;32(11):e4792. doi:10.1002/PRO.4792

19. Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM. NPDock: a web server for protein–nucleic acid docking. *Nucleic Acids Res*. 2015;43(Web Server issue):W425. doi:10.1093/NAR/GKV493

20. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol*. 2007;372(3):774-797. doi:10.1016/J.JMB.2007.05.022

21. Delano W. The PyMOL Molecular Graphics System. Published online 2002.

22. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701-1718. doi:10.1002/JCC.20291

23. Lindahl E, Hess B, Spoel D van der. GROMACS 3.0: a package for molecular simulation and trajectory analysis: a package for molecular simulation and trajectory analysis. *J Mol Model*. 2001;7(8):306-317. Accessed January 4, 2025. https://research.rug.nl/en/publications/gromacs-30-a-package-for-molecular-simulation-and-trajectory-anal

24. Brysbaert G, Lensink MF. Centrality Measures in Residue Interaction Networks to Highlight Amino Acids in Protein-Protein Binding. *Front Bioinforma*. 2021;1. doi:10.3389/FBINF.2021.684970

25. Dunn R, Dudbridge F, Sanderson CM. The use of edge-betweenness clustering to investigate biological function in protein interaction networks. *BMC Bioinformatics*. 2005;6(1):1-14. doi:10.1186/1471-2105-6-39/TABLES/13

26. Cirino PC, Mayer KM, Umeno D. Generating mutant libraries using error-prone PCR. *Methods Mol Biol*. 2003;231:3-10. doi:10.1385/1-59259-395-X:3

27. Deppmann CD, Acharya A, Rishi V, et al. Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana : a comparison to Homo sapiens B-ZIP motifs. *Nucleic Acids Res*. 2004;32(11):3435-3445. doi:10.1093/NAR/GKH653

28. Grigoryan G, Keating AE. Structure-based Prediction of bZIP Partnering Specificity. *J*

Mol Biol. 2006;355(5):1125-1142. doi:10.1016/j.jmb.2005.11.036

29. Li M, Lin W, Hinckley W, et al. Uncovering the "ZIP code" for bZIP dimers reveals novel motifs, regulatory rules and one billion years of cis-element evolution. *bioRxiv*. Published online April 17, 2022:2022.04.17.488518. doi:10.1101/2022.04.17.488518

30. Negre CFA, Morzan UN, Hendrickson HP, et al. Eigenvector centrality for characterization of protein allosteric pathways. *Proc Natl Acad Sci U S A*. 2018;115(52):E12201-E12208. doi:10.1073/PNAS.1810452115

31. Vinson CR, Hai T, Boyd SM. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: Prediction and rational design. *Genes Dev*. 1993;7(6):1047-1058. doi:10.1101/gad.7.6.1047

32. Krylov D, Mikhailenko I, Vinson C. A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J*. 1994;13(12):2849-2861. doi:10.1002/J.1460-2075.1994.TB06579.X

33. Nakajima H, Kobayashi I, Adachi Y, Hisatomi O. Transmission of light signals from the light-oxygen-voltage core via the hydrophobic region of the β-sheet surface in aureochrome-1. *Sci Reports 2021 111*. 2021;11(1):1-10. doi:10.1038/s41598-021-91497-5

34. Tzamarias D, Pu WP, Struhl K. Mutations in the bZIP domain of yeast GCN4 that alter DNA-binding specificity. *Proc Natl Acad Sci*. 1992;89(6):2007-2011. doi:10.1073/PNAS.89.6.2007

35. Kroth PG, Wilhelm C, Kottke T. An update on aureochromes_ Phylogeny – mechanism – function. *J Plant Physiol*. 2017;217:20-26. doi:10.1016/j.jplph.2017.06.010

36. Huang Y jiang, Wang L, Zheng M gang, Zheng L, Tong Y li, Li Y. Overexpression of NgAUREO1, the gene coding for aurechrome 1 from Nannochloropsis gaditana, into Saccharomyces cerevisiae leads to a 1.6-fold increase in lipid accumulation. *Biotechnol Lett*. 2014;36(3):575-579. doi:10.1007/s10529-013-1386-0

37. Mann M, Serif M, Wrobel T, et al. The Aureochrome Photoreceptor PtAUREO1a Is a Highly Effective Blue Light Switch in Diatoms. *iScience*. 2020;23(11). doi:10.1016/J.ISCI.2020.101730

38. Huysman MJJ, Fortunato AE, Matthijs M, et al. AUREOCHROME1a-Mediated Induction of the Diatom-Specific Cyclin dsCYC2 Controls the Onset of Cell Division in Diatoms (Phaeodactylum tricornutum). *Plant Cell*. 2013;25(1):215-228. doi:10.1105/tpc.112.106377

39. Wang Y, Xue P, Cao M, Yu T, Lane ST, Zhao H. Directed Evolution: Methodologies and Applications. *Chem Rev*. 2021;121(20):12384-12444. doi:10.1021/ACS.CHEMREV.1C00260

40. Chen K, Arnold FH. Engineering new catalytic activities in enzymes. *Nat Catal*. 2020;3(3):203-213. doi:10.1038/S41929-019-0385-5;SUBJMETA

PLAGIARISM CERTIFICATE

DrillBit

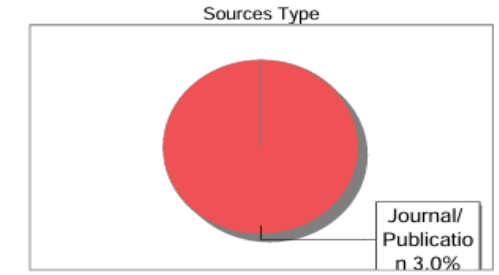
The Report is Generated by DrillBit Plagiarism Detection Software

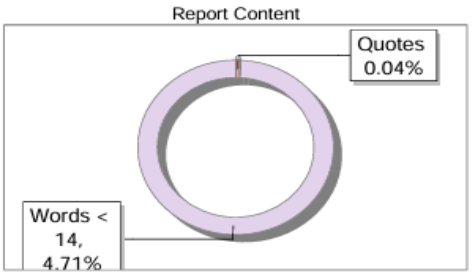
Submission Information

Author Name	Madhurima Khamaru
Title	Characterization of Aureochromes: the Blue Light Responsive LOV Photoreceptor-Transcription Factors from Photosynthetic Marine Algae
Paper/Submission ID	4516775
Submitted by	devrani.dbs@presiuniv.ac.in
Submission Date	2025-10-13 00:39:30
Total Pages, Total Words	15, 5138
Document type	Synopsis

Result Information

Similarity **3 %**





Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

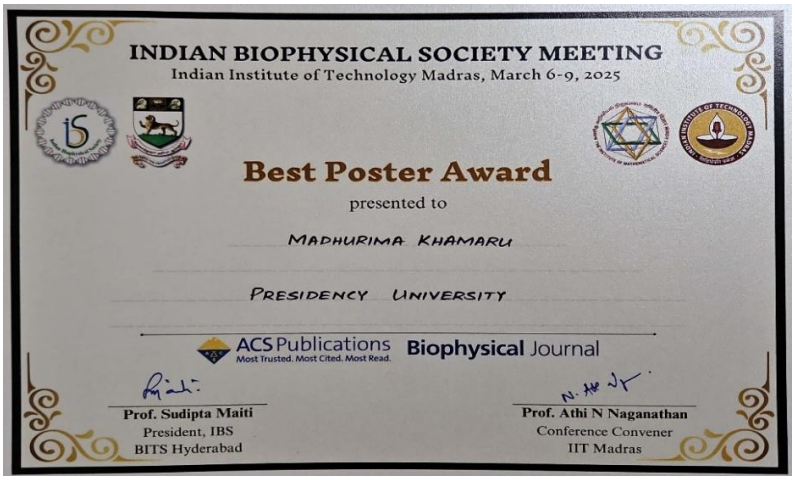
A Unique QR Code use to View/Download/Share Pdf File

ADDITIONAL DOCUMENTS-

A. CERTIFICATES OF CONFERENCES ATTENDED

Certificate of Participation
This is to certify that
Madhurima Khamaru
Presidency University, Kolkata, India
was awarded 3rd Prize in POSTER presentation at the two day International Conference (Online) on
"Biomolecules To Biome" held on August 24-25, 2022
Organized by Department of Life Sciences, Presidency University, Kolkata, India.
Dr. Ranjana Pal & Dr. Puja Ray
Joint Organizing Secretaries
Biomolecules To Biome - 2022
Prof. Prabir Kumar Mukhopadhyay
HOD (Life Sciences)
Presidency University, Kolkata
(Chairman)
Prof. Sankar Bose
Dean of Faculty of Natural and
Mathematical Sciences
(Patron)

International Meeting
4th IABS 2023
An interdisciplinary Approach to Biological Sciences
Certificate of Poster Presentation/ Participation
is awarded to
Madhurima Khamaru
for attending the 4th IABS 2023 International Meeting on
An interdisciplinary Approach to Biological Sciences
from February 1st-3rd, 2023
at the Indian Association for the Cultivation of Science, Kolkata
Convener
Co-Convener



B. FIRST PAGE OF PUBLISHED MANUSCRIPT

Cells
Tissues
Organs

Research Article

Cells Tissues Organs 2024;213:133–146
DOI: 10.1159/000527593

Received: June 6, 2022
Accepted: October 12, 2022
Published online: October 19, 2022

Assessing Combinatorial Diversity of Aureochrome Basic Leucine Zippers through Genome-Wide Screening

Madhurima Khamaru[a] Deep Nath[b] Devrani Mitra[a] Soumen Roy[b]

[a]Department of Life Sciences, Presidency University, Kolkata, India; [b]Department of Physics, Bose Institute, Kolkata, India

Downloaded from http://karger.com/cto/article-pdf/213/2/133/4188358/000527593.pdf by University of Edinburgh user on 12 January 2025

Keywords
Aureochrome · Basic leucine zipper
Homo/heterodimerization · Dimer compatibility and stability · Complex networks · Information theory

Abstract
Aureochromes are unique blue light-responsive light-oxygen-voltage (LOV) photoreceptors cum basic leucine zipper (bZIP) transcription factors (TFs), present exclusively in photosynthetic marine stramenopiles. Considering the availability of the complete genome sequence, this study focuses on aureochromes from *Ectocarpus siliculosus*. Aureochromes mediate light-regulated developmental responses in these brown photosynthetic algae. Both the LOV sensor and the bZIP effector show overall sequence-structure conservation. The structurally similar LOV + bZIP modules of aureochrome homologs/paralogs prefer a dimeric state. Besides a heterogeneous linker connecting the sensor-effector and a flexible N-terminal region, the sequence composition of both domains is vital. Aureochromes execute diverse cellular responses in different photosynthetic stramenopiles – though their activities can vary even within a given algal species. Therefore, it is important to understand whether aureochromes select dimerization partners from the same family or interact with other bZIPs as well. To regulate multifarious biological activities, it is possible that aureochromes activate the global TF interaction network. Following homo/heterodimer modeling, we address the compatibility of dimerization partners by screening through heptad repeats. We evaluate the dimer interface area in terms of gain in solvation energy and the number of hydrogen bonds/salt bridge interactions. We further explore the relative stability of these structures from a graph-theoretic perspective through well-studied measures such as the energy of the graph, average participation coefficient, and betweenness centrality. Furthermore, we also conduct an information-theoretic analysis using hitherto understudied measures such as network information centrality and Kullback-Leibler divergence. We find that all our investigations into the relative stability of the dimers using diverse methods from bioinformatics, network science, and information theory are in harmonious agreement. Coupling preferences of monomers in aureochromes can be further translated to design novel optogenetic tools useful for understanding human development and disease.
© 2022 S. Karger AG, Basel

Introduction

The intrinsic cellular processes governing life are regulated by transcription factors (TFs). They bridge the gene regulation pattern of cells with diverse extracellular or intracellular signaling cascades [Weidemüller et al., 2021].

karger@karger.com
www.karger.com/cto
© 2022 S. Karger AG, Basel

Correspondence to:
Devrani Mitra, devrani.dbs@presiuniv.ac.in
Soumen Roy, soumen@jcbose.ac.in

[Book Chapter]

Designing synthetic transcription factors: A structural perspective

Aparna Boral†, Madhurima Khamaru†, and Devrani Mitra*
Department of Life Sciences, Presidency University, Kolkata, West Bengal, India
*Corresponding author: e-mail address: devrani.dbs@presiuniv.ac.in

Abstract

In this chapter, we discuss different design strategies of synthetic proteins, especially synthetic transcription factors. Design and engineering of synthetic transcription factors is particularly relevant for the need-based manipulation of gene expression. With recent advances in structural biology techniques and with the emergence of other precision biochemical/physical tools, accurate knowledge on structure-function relations is increasingly becoming available. Besides discussing the underlying principles of design, we go through individual cases, especially those involving four major groups of transcription factors—basic leucine zippers, zinc fingers, helix-turn-helix and homeodomains. We further discuss how synthetic biology can come together with structural biology to alter the genetic blueprint of life.

Abbreviations

bZIP	basic leucine zipper
HTH	helix-turn-helix
STF	synthetic transcription factor
ZnF	zinc finger

† Equal contribution.

Advances in Protein Chemistry and Structural Biology
ISSN 1876-1623
https://doi.org/10.1016/bs.apcsb.2021.12.003

Copyright © 2022 Elsevier Inc.
All rights reserved.

ACKNOWLEDGEMENT

I am indebted to my supervisor Dr. Devrani Mitra for her invaluable support and guidance throughout the tenure of my research work. I offer my sincere thanks to Dr. Soumen Roy for his academic insights and guidance.

I am profoundly grateful to the entire Department of Life-Sciences, Presidency University for providing excellent infrastructure, instrumental facilities and conducive environment for research.

I am thankful to the faculties of Department of Life-Sciences, Presidency University whose dedication to teaching significantly shaped my understanding of the subject of my research.

I would like to express my deepest gratitude to the RAC members Dr. Malay Das and Dr. Sanghamitra Dey for their insightful feedback and constructive suggestions that substantially improved the quality of work.

I extend my thanks to Dr. Aparna Mukhopadhyay, the members of PhD committee, HoD of Life-Sciences and the external experts for their efforts in critical assessments and seamless conduction of activities required for smooth running of the PhD program.

I am thankful to Debarshi Bose, Pracheta Pal, Sayani Sikdar, Aniket Laskar, Saptarshi Pal, Deep Nath, Soumyajit, Esha, Abhilasha, Dwaipayan, Joyeeta, Joykesh, Sujata, Payel, Asim, Aparna, Suchayan and Anwesha di.

Last but not the least, I am extremely grateful to my family and funding agency without whom it would not have been possible.